

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

5-2004

The role of the toll-like receptor pathway in susceptibility to inflammatory bowel disease.

Nigel Crawford 1974-
University of Louisville

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Recommended Citation

Crawford, Nigel 1974-, "The role of the toll-like receptor pathway in susceptibility to inflammatory bowel disease." (2004). *Electronic Theses and Dissertations*. Paper 288.
<https://doi.org/10.18297/etd/288>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

**THE ROLE OF THE TOLL-LIKE RECEPTOR PATHWAY IN
SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE**

By

Nigel Crawford

B.Sc. (Hons), University of Liverpool, 1996

M.B. Ch.B., University of Liverpool, 1998

A Dissertation

Submitted to the Faculty of the
Graduate School of the University of Louisville
In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

Department of Physiology and Biophysics
University of Louisville
Louisville, Kentucky

May 2004

**THE ROLE OF THE TOLL-LIKE RECEPTOR PATHWAY IN
SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE**

By

Nigel Crawford

B.Sc. (Hons), University of Liverpool, 1996

M.B. Ch.B., University of Liverpool, 1998

A Dissertation Approved on

March 12th, 2004

Date

by the following Dissertation Committee:

Susan Galandiuk, M.D., Dissertation Director

Irving G. Joshua, Ph.D., Co-Advisor

Gary A. Cobbs, Ph.D.

Jeff C. Falcone, Ph.D.

John T. Fleming, Ph.D.

DEDICATION

This dissertation is dedicated to my wife

Sarah

Without her support, love and encouragement none of this would have been possible.

ACKNOWLEDGEMENTS

I owe an immense debt of gratitude to my mentor, Dr Susan Galandiuk, who invited me to Louisville to become the Price Fellow in Surgical Research in February 2001. I am extremely grateful to her for her wise advice, patience and mentorship. Likewise, I'd like to thank our lab manager, Robert Eichenberger, without whose help and technical expertise this project would never have made it off the ground. I would also like to acknowledge the generosity of the Price family for providing the funding for my position.

I would like to express my sincere appreciation to the members of my Ph.D. committee from the Department of Physiology and Biophysics. Drs Joshua, Falcone and Fleming have helped shape this dissertation in an extremely positive manner, and their advice has proven invaluable. I also wish to express my thanks to the final member of my Ph.D. committee, Dr Gary Cobbs, for helping with the most complex portion of this dissertation – the statistical analysis.

I want to express my sincere gratitude to Dr Hiram Polk for his constructive criticism of my work and help in applying a logical approach to data presentation and to scientific research in general.

I would like to thank Mark Doll and Dr David Hein from the Department of Pharmacology and Toxicology in the University of Louisville for their assistance with primer and probe design, and use of their equipment in SNP-specific PCR experiments.

I'd like to express thanks to everyone who's helped me with the seemingly never-ending glut of PCR's. These people would include Alisa Funke, Valentina Kolodko, Scott Kelley, Brooke Grant, Michael Young and Daniel Colliver.

Finally, I'm immensely grateful to my family, especially my parents Anne and David Crawford for their encouragement and support during my entire time I have been living outside of the United Kingdom.

ABSTRACT

THE ROLE OF THE TOLL-LIKE RECEPTOR PATHWAY IN SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE

Nigel Crawford, B.Sc. (Hons), M.B. Ch.B.

February 16th, 2004

Inflammatory bowel disease (IBD) is a chronic autoimmune disorder that is subdivided into Crohn's disease (CD), ulcerative colitis (UC), and indeterminate colitis (IC). Epidemiological studies have proven that genetic variation increases susceptibility to IBD, with multiple abnormal genes combined with environmental factors being responsible for disease development. Characterization of these susceptibility genes remains of critical importance to improving understanding of IBD pathogenesis.

The overall aim of this study is to discover new IBD susceptibility genes. The initial approach was to study a number of previously described IBD susceptibility loci through characterization of peak LOD score short tandem repeat markers using population- and family-based methods. The IBD1, IBD2, and IBD5 loci were shown to be associated with different forms of IBD in the study cohort. This work formed the basis of the next series of experiments, where the IBD2 locus was mapped in detailed. IBD2 was shown to be associated with IC, colonic CD and to some extent UC.

Focus of this study then turned to two candidate genes: interleukin-1 receptor-associated kinase-M (*IRAK-M*), an inhibitor of the strongly proinflammatory Toll-like

receptor pathway, and myeloid differentiation factor 88 (*MyD88*), an adaptor protein involved in macrophage apoptosis. These genes were selected on the basis of their biological plausibility and genomic location (within IBD2 and IBD9 respectively). The first phase of this study was to define candidate gene colonic mucosal expression levels in IBD, and to screen coding and regulatory regions for polymorphisms. Both genes were significantly over-expressed in IBD-affected mucosa, and contained a total of seven commonly occurring single nucleotide polymorphisms (SNPs).

These polymorphisms were then characterized in the entire population to define disease-associations. SNPs within the 3'-untranslated region of *MyD88* were associated with late onset CD in patients with a family history of IBD. However, a number of concerns do exist regarding the low statistical power of these analyses. In contrast, *IRAK-M* is unlikely to play a role in IBD susceptibility in this population.

In conclusion, preliminary evidence is presented to show that *MyD88* may be an IBD susceptibility gene. Its precise role in CD development will be clarified in further studies.

TABLE OF CONTENTS

	Page
ABSTRACT.....	vi
LIST OF FIGURES	xii
LIST OF TABLES.....	xiii
CHAPTER	
I. INTRODUCTION	1
II. GENERAL BACKGROUND	5
Introduction	5
IBD Genetics	7
<i>NOD2</i> : The IBD Susceptibility Gene.....	10
Gastrointestinal Immunological Dysfunction and Inflammatory Mediators in IBD.....	11
Innate Immunity	14
Toll-Like Receptors	15
TLR Family Signaling Pathways	18
Interleukin-1 Receptor-Associated Kinases.....	22
IRAK-M: An Inhibitor of TLR Signaling.....	24
Mechanism of Action of IRAK-M.....	25
MyD88 and TLR-Mediated Apoptosis	28

III.	FUNDAMENTAL QUESTIONS.....	33
	Proposed Concepts	34
	Hypothesis #1	35
	Hypothesis #2.....	37
	Hypothesis #3.....	37
	Hypothesis #4.....	40
	Biological Plausibility and Candidate Gene Selection.....	42
IV.	MATERIALS AND METHODS	44
	Patient Population and Study Participation.....	44
	Characterization of IBD Susceptibility Loci STR Markers	46
	Characterization of TLR Pathway IBD Candidate Genes.....	50
V.	CHARACTERIZATION OF GENOTYPE-PHENOTYPE RELATIONSHIPS FOR INFLAMMATORY BOWEL DISEASE SUSCEPTIBILITY LOCI USING MULTIPLE SHORT TANDEM REPEAT GENETIC MARKERS.....	64
	Overview	64
	Introduction	65
	Materials and Methods.....	66
	Results	72
	Discussion	77
VI.	CHARACTERIZATION OF GENOTYPE-PHENOTYPE RELATIONSHIPS WITHIN THE INFLAMMATORY BOWEL DISEASE SUSCEPTIBILITY LOCUS IBD2	83

	Overview	83
	Introduction	84
	Materials and Methods	85
	Results	90
	Discussion	101
VII.	PRELIMINARY ASSESSMENT OF INTERLEUKIN-1 RECEPTOR ASSOCIATED KINASE-M AND MYELOID DIFFERENTIATION FACTOR 88 AS INFLAMMATORY BOWEL DISEASE CANDIDATE GENES	106
	Overview	106
	Introduction	107
	Materials and Methods	110
	Results	119
	Discussion	132
VIII.	IS THERE A ROLE FOR <i>IRAK-M</i> AND <i>MYD88</i> IN SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE?	138
	Overview	138
	Introduction	139
	Materials and Methods	141
	Results	146
	Discussion	160
IX.	CONCLUDING REMARKS	167
	Summary of Findings	167

	Directions of Future Research.....	170
X.	REFERENCES	173
XI.	CURRICULUM VITAE.....	183

LIST OF FIGURES

FIGURE

1.	Ligand specificities for different TLRs.....	17
2.	The interleukin-1 receptor (IL-1R)/Toll-like receptor (TLR) signaling pathway.	19
3.	Proposed mechanism of action of IRAK-M.....	27
4.	TLR-activated inflammatory and apoptotic pathways bifurcate at the level of MyD88 in macrophages.	31
5.	Example of a polymorphic and non-polymorphic dHPLC chromatogram used for screening candidate gene PCR amplicons for polymorphisms.	55
6.	Post-PCR determination of fluorescence from SNP-specific PCR reactions – typical read-out.	61
7.	Polymorphisms within the 527bp <i>IRAK-M</i> exon 2 amplicon.	123
8.	Polymorphism within the 705bp <i>IRAK-M</i> exon 5 amplicon.....	125
9.	Polymorphism within the 769bp amplicon containing the 3'-end of <i>IRAK-M</i> exon 12 (exon 12.2).	127
10.	Polymorphism within the 705bp <i>MyD88</i> exon 3 amplicon.	129
11.	Polymorphisms within the 816bp amplicon containing the 3'-end of the 3'-untranslated region of <i>MyD88</i> (3'-UTR3).....	131
12.	Molecular structure of valine and isoleucine.	135

LIST OF TABLES

TABLE

1.	STR markers utilized in the IBD susceptibility loci study.....	70
2.	STR allele frequencies for IBD1, IBD2, and IBD5	73
3.	Case-control association test global P-values for IBD loci peak LOD score STRs.	74
4.	Allele frequencies for the short tandem repeats (STR) D12S83, D12S1056, and D12S1022.....	92
5.	Genotype frequencies for the short tandem repeats (STRs) D12S83, D12S1056, and D12S1022.....	94
6.	Case-control association global P-values for disease group versus control comparisons for IBD2 STRs.....	95
7.	Transmission disequilibrium test analysis of D12S83 alleles among indeterminate colitis families.	98
8.	Pedigree disequilibrium test analysis of D12S1056 alleles among ulcerative colitis families.	99
9.	Haplotypes for the short tandem repeat markers D12S1056 and D12S83 in the indeterminate colitis population.....	100
10.	PCR amplicons designed for <i>IRAK-M</i> using VectorNTI software.	114

11.	PCR amplicons designed for <i>MyD88</i> using VectorNTI software.....	116
12.	Polymorphic candidate gene amplicons on dHPLC analysis of the screening population.	120
13.	Demographic data for IBD patients in candidate gene study.....	142
14.	Disease characteristics of the CD population used in candidate gene studies. ...	143
15.	Primers and fluorescently-labeled oligonucleotide probes used in candidate gene SNP-specific PCR experiments.....	145
16.	<i>IRAK-M</i> and <i>MyD88</i> genotype and allele frequencies for patients and controls.	149
17.	<i>IRAK-M</i> mutant allele frequencies for CD and UC populations at the sub-phenotype level.	152
18.	<i>MyD88</i> mutant allele frequencies for CD and UC populations at the sub-phenotype level.	154
19.	Estimated haplotype frequencies for <i>MyD88</i> SNPs in early and late onset CD populations (Vienna Classification groups A1 and A2).	156
20.	Significant Pedigree Disequilibrium Test analyses for <i>MyD88</i> sub-phenotypic groups.	159
21.	<i>MyD88</i> haplotype frequencies for families where the proband has UC without the presence of extra-intestinal manifestations.	160

CHAPTER I

INTRODUCTION

Almost all human disorders are affected to some extent by genetic factors. Such hereditary influences will tend to predispose to the development of either individual diseases (e.g. rheumatoid arthritis) or specific disease types (e.g. autoimmune disease). Consideration of such factors allows diseases to be broadly classified into 'Mendelian' or 'complex' based on their inheritance pattern.

Mendelian or single-gene traits are directly related to distinct genetic or biochemical defects. This type of inheritance pattern is rare and is only apparent in a small number of human disorders. Complicating factors, such as incomplete penetrance (that is, not all genetically predisposed individuals will develop the disease) and variable age of onset are often evident in this type of disease, but all show basic Mendelian inheritance patterns (1).

In contrast to Mendelian disorders, complex diseases are highly prevalent within human populations. Common examples include essential hypertension, osteoarthritis and schizophrenia. These types of disorders differ in that the influence of genetic factors is much less well defined. Overall, the influence of hereditary factors appears to be much more subtle, and rather than being overtly involved in pathogenic processes as is the case with Mendelian disorders, genetic variation within a population causes susceptibility to a wide variety of different disorders.

Genetic predisposition to complex disease development is likely caused by the presence of multiple variant or abnormal genes within an individual's genome. The effect of such abnormalities may be small in isolation, and an individual possessing a single variant copy of such a gene may lead a perfectly normal existence. However, it is the combination of variant genes within a genome, and how such genes interact both with each other and disease-specific environmental stimuli that lead to the development of a particular disorder. Differential rates of complex diseases in monozygotic as compared to dizygotic twins, as well as a higher frequency of disease in first-degree relatives of affected individuals when compared to the general population, have conclusively proven that genetic, as well as environmental factors, contribute to complex disease pathogenesis (1).

Considerable effort has been made to define the genetic basis of many Mendelian disorders, with a number of approaches being commonly used to elucidate disease-causing genes. One of the most frequently employed techniques is the candidate gene approach, where the biochemical nature of a disease is to some extent understood. Candidate gene studies have proven successful in a number of instances. For example, a small subset of type 2 diabetes has been shown to be caused by mutation of genes encoding insulin and the insulin receptor (1). An alternative method for defining disease-causing genes is positional cloning. This type of approach is typically used when the biochemical nature of a disease is unknown. This method was successfully applied in the cloning of the cystic fibrosis gene on chromosome 7, as well as the Huntington's disease gene on chromosome 4 (2). It must be noted that relatively few genes have been

identified using either of these approaches, and much work remains in defining those genetic abnormalities underlying other Mendelian disorders.

Similar approaches can be employed to define genetic abnormalities in complex diseases. However, despite extensive efforts, relatively few genetic regions and even fewer genes have been positively implicated in complex disease pathogenesis (1). Some of the most successful examples include mapping of complex disease susceptibility to the human leukocyte antigen (HLA) region of chromosome 6, which has been shown to be an etiological factor in many human disorders (3). Examples of HLA-associated diseases include a number of inflammatory diseases like rheumatoid arthritis (4), ankylosing spondylitis and inflammatory bowel disease (5). Despite these successes, limited progress has been made in elucidating non-HLA genes involved in complex diseases pathogenesis. The outcomes of on-going studies in a variety of diseases are eagerly anticipated in many quarters since knowledge of the genetic basis of complex disease susceptibility is likely to lead to a greater overall understanding of the numerous biological steps and interactions that must occur for an individual to develop a disease. Such knowledge may have a profound impact on how complex diseases are managed in the clinical setting. For instance, knowledge of specific disease-causing genetic abnormalities may lead to development of novel therapies that may alleviate symptoms. Additionally, it may become possible to define an individual's susceptibility to developing a specific disease, thus enabling such an individual to minimize exposure to environmental factors involved in the pathogenesis of the disorder.

The focus of this dissertation will be inflammatory bowel disease (IBD). IBD is a debilitating chronic inflammatory autoimmune disorder of the gastrointestinal tract. It has

a relatively high frequency, affecting approximately one million Americans (6).

Symptoms tend to first occur in early adulthood and are often debilitating, with surgery frequently being required to provide alleviation. Additionally, patients have a significantly increased risk of developing colorectal cancer (7).

IBD is known to have a complex, non-Mendelian form of inheritance. The overall aim of this dissertation is to define IBD susceptibility genes, firstly by defining genomic regions of susceptibility in the study population, and secondly by performing candidate gene studies. The nature of IBD and the candidate gene pathway that will be the focus of this dissertation will be considered first.

CHAPTER II

GENERAL BACKGROUND

Introduction

Inflammatory bowel disease (IBD) encompasses a number of chronic inflammatory disorders of the gastrointestinal tract that share a variety of clinical characteristics, the most notable of which are profuse diarrhea and abdominal pain. Symptoms are frequently debilitating, and multiple surgical procedures may be required if medical treatment fails. Patients are generally sub-classified as suffering from either Crohn's disease (CD) or ulcerative colitis (UC) on the basis of clinical, radiological and histological grounds. However, approximately 15% of colonic IBD cases display ambiguous disease features, and a definitive diagnosis of either UC or CD cannot be made. In such instances, individuals are diagnosed as suffering from 'indeterminate colitis' (IC) (8). Inflammation in UC and IC is limited to the colon and rectum, whereas inflammation in CD, although most frequently affecting the terminal ileum, can involve any part of the gastrointestinal tract. Additionally, all forms of IBD may be associated with extra-intestinal manifestations such as arthritis, cholangitis, iritis and dermatological problems.

It is estimated that IBD affects up to one million Americans (6), with the prevalence of UC being slightly higher than that of CD. On a global scale, IBD is predominantly found in areas of the developed world (historically Scandinavia, the

United Kingdom, and the United States), although the mid-1980s and 1990s did see a dramatic increase in incidence in continental Europe, the Middle East, the Pacific Rim, Africa and Latin America (9). In general, however, IBD is considered to be a disease of urban areas, with its prevalence typically being low in rural regions. Epidemiological studies have shown that IBD tends to develop in early adulthood: CD is diagnosed most frequently in individuals in their 20's and UC in individuals in their 30's. Approximately 10% to 15% of individuals are diagnosed as suffering with IBD before adulthood (9). The sex-distribution of this disease is also different, with all forms of IBD being slightly more prevalent in females (9).

It is widely accepted that IBD, like all complex non-Mendelian disorders, develops in a genetically susceptible individual in response to exposure to environmental stimuli (10). Despite the fact that both genetic and environmental factors have been implicated in the pathogenesis of IBD (11), the precise mechanisms of action of such factors, and more importantly, how environmental and genetic determinants interact to produce specific IBD phenotypes remain largely unknown. Similarly, the pathologic mechanism underlying the development of the chronic inflammation seen in IBD continues to be elusive. Many models of IBD pathogenesis have been suggested, although none has yet been proven correct. One of the most persuasive models postulates that IBD develops as a consequence of an abnormal host response to normal intestinal bacterial flora (12).

Environmental factors implicated in IBD pathogenesis are diverse, and are for the most part associated with Western lifestyles of the second half of the twentieth century. Environmental factors that have been proposed as increasing the risk of developing IBD

include cigarette smoking, oral contraceptive use, breast-feeding and the measles vaccine to name but a few (13). The most heavily studied of such environmental factors is cigarette smoking, which has been shown to have opposite effects on disease activity in UC and CD. Symptoms of CD are exacerbated by exposure to tobacco, whereas symptoms of UC are actually alleviated (14). Although, elucidation of such environmental stimuli is of interest, such studies do have one notable drawback i.e. they do not provide a clear mechanism of action as to how the specific factor(s) leads to IBD development.

IBD Genetics

Numerous epidemiological studies have conclusively demonstrated that IBD pathogenesis has a prominent genetic component and that certain individuals are more prone to developing IBD. Twin concordance rates of IBD have been used to estimate the relative effect of genetic factors in IBD pathogenesis and have proven a powerful tool in determining the relative roles of genetic and environmental influences in IBD. Such studies have demonstrated a greater concordance of IBD in monozygotic twin pairs compared to a dizygotic twin pairs (11;15). Additionally, these effects appear to be more prominent in CD than in UC. Assuming that each twin-pair has been subject to the same environmental stimuli, it must be concluded that genetic factors are responsible for this difference in disease concordance.

The findings of twin studies have been augmented by observations of ethnic differences in disease frequency. Such studies have illustrated that IBD is primarily a disease of Caucasians (9), with its prevalence being highest in Jewish populations (16).

Additional compelling evidence is provided by familial aggregation studies, which have estimated that between 6% and 32% of patients with IBD have an affected first- or second-degree relative (17).

Consideration of both genetic and environmental factors has led many to postulate that the classification of IBD into UC and CD is too narrow. Many genetic studies have led researchers to suggest that IBD actually represents a heterogeneous group of inflammatory disorders, and that different clinical subgroups are caused by different genetic defects. Exposure to environmental stimuli is, however, required for the disease process to become active (17). An individual with a specific IBD phenotype will probably carry some susceptibility genes that are common to all types of IBD, and a number of others that confer susceptibility to a specific IBD sub-phenotype. This assumption has been supported by genetic linkage studies, which are discussed in greater detail below.

Genome-wide scans have been a particularly useful tool in defining genetic regions that confer increased susceptibility to developing IBD. These studies employ short tandem repeat (STR) markers, which are a type of variable-number tandem repeat marker, consisting of 2-6 nucleotides, repeated a variable number of times. STRs have proven extremely useful tools for a number of reasons: first, they show large variation within any population; second, they are widely dispersed throughout the genome; and finally, if an STR is in close physical proximity to a disease gene, meiotic recombination between the marker and disease gene is highly unlikely to occur, meaning that a specific STR allele will be co-inherited with the disease gene. Therefore, markers closest to the disease gene would show the strongest correlation with disease patterns. Hundreds of

STRs are employed in genome-wide scans. The outcome of these scans has been to define genomic regions with a high degree of STR allele sharing among disease-affected sibling-pairs. In other words, IBD-affected sibling-pairs share STR marker alleles that are in close physical and genetic proximity to the same (shared) disease gene. Genomic regions that demonstrate a high frequency of STR allele sharing are termed 'susceptibility loci'. A number of such susceptibility loci have been defined in IBD, the most well characterized of which have been termed IBD1-9 (18-25). These loci tend to confer susceptibility to different IBD phenotypes. For example, IBD1 has been demonstrated to be important in susceptibility primarily CD, whereas IBD2 confers susceptibility to both UC and CD.

There are a number of obvious drawbacks when one defines the heritable component of a disease in terms of susceptibility loci. The most notable of these is that although these loci do represent a genetic region of susceptibility, they do not give any information as to the nature of the susceptibility gene contained within. The overall aim of studying the genetics of complex disease must be to elucidate susceptibility genes. Therefore, susceptibility loci only serve to direct researchers to genomic regions that will contain susceptibility gene(s). Another important problem with susceptibility loci is that they are very large in both physical and genetic terms. They span many millions of base pairs and contain hundreds of genes. Finding the one disease-causing gene within such a large genetic region therefore presents considerable challenges.

***NOD2*: The IBD Susceptibility Gene**

Only one gene has thus far been conclusively implicated in IBD susceptibility (26;27). This gene, nucleotide oligomerization domain 2 (*NOD2*), is located on chromosome 16q12 and is likely to account for the majority, but not all, of the linkage observed within the IBD1 susceptibility locus (24). The biological function of *NOD2* has been a matter of great interest since the discovery of its role in CD susceptibility, and strong arguments exist to support its role in both up-regulation of the pro-inflammatory transcription factor NF- κ B and induction of apoptosis in macrophages (28). It is known that expression of *NOD2* is limited to monocytes and dendritic cells, and that it acts as an intracellular pattern recognition receptor for bacterial peptidoglycan (28). Stimulation of *NOD2* leads to NF- κ B up-regulation and apoptosis. *NOD2* disease-causing mutations (DCMs) have been shown to interfere with the ligand recognition process and results in impaired *NOD2*-mediated NF- κ B activation. This finding presents somewhat of a paradox given that CD is associated with increased NF- κ B activity. Further work is therefore required to clarify the mechanism by which *NOD2* mutations cause susceptibility to CD.

Three independent DCMs within *NOD2* (one frame-shift mutation [1007fs] and two missense mutations [R702W and G908R]) have been shown to be associated with CD. These three mutations are estimated to be carried by 25-35% of CD patients and 7-15% healthy controls (29). The relative risk for CD development has been estimated as approximately 2-3 in individuals carrying one DCM and 20-40 in individuals who are homozygous mutants or compound heterozygotes (30;31).

Investigation of *NOD2* mutations in different CD phenotypic subgroups has proved extremely worthwhile. *NOD2* mutations have been shown to be associated with inflammatory processes localized to the terminal ileum, fibrostenotic disease behavior and early age of onset (31-36). Additionally, carriage of these mutations only appears to be of significance in Caucasian populations, with *NOD2* not being involved in susceptibility to CD in Chinese (37) and Japanese populations (38).

Gastrointestinal Immunological Dysfunction and Inflammatory Mediators in IBD

There is accumulating evidence to suggest that both UC and CD develop as a consequence of an abnormal host response to normal intestinal bacteria (12;39). Rodent models of IBD have demonstrated that the presence of luminal flora is necessary for the development of intestinal inflammation. Mutant rat and mouse strains prone to developing “spontaneous colitis” do not develop intestinal inflammation when bred in a germ-free environment (40). The nature of defects underlying development of inflammation in IBD remains unclear. It is probable that inflammation either develops because of intrinsic defects in gastrointestinal immunoregulation or defects in bowel mucosal barrier function, leading to continued immune effector cell stimulation. There is compelling evidence for both theories. Monocytes within the lamina propria of individuals with CD are spontaneously activated, possibly by gram-negative bacteria (41). Such spontaneous activation of phagocytes can directly cause significant alterations in barrier function (41).

Mixed leukocytic infiltration of gastrointestinal mucosa is a prominent feature of IBD. CD4⁺ T-cells are the most common form of activated mononuclear cells infiltrating the gut (42). Enhanced adhesion molecule and chemokine expression in inflamed IBD tissue promotes recruitment of these cells to the intestinal mucosa. These CD4⁺ T-cells play an important role in the pathogenesis of tissue damage in IBD, especially in CD (42-44). Different T-cell subsets appear to predominate in CD and UC, with Th1 T-cells being more common in inflamed CD mucosa, and atypical Th2 T-cells being more common in inflamed UC mucosa (39).

Both CD and UC are characterized by infiltration of the bowel mucosa by activated macrophages, in addition to this lymphocytic infiltrate. It is believed that this occurs as a consequence of lymphocyte Th1 cytokine production in CD (e.g. interferon-gamma [IFN γ] and IL-2), although how this cell population becomes active in UC is less certain. Activated macrophages, along with other mucosal cell types, produce a wide range of pro-inflammatory cytokines, including TNF α , IL-1 β , IL-6, IL-8, and IL-16. Production of vast quantities of such broadly active cytokines, together with enhanced expression of a wide variety of inflammatory mediators, exacerbates the inflammatory process, eventually leading to the clinical manifestations of disease (39). Other pro-inflammatory agents include chemokines, arachidonic acid metabolites (e.g. prostaglandins and leukotrienes) and reactive oxygen metabolites (43;45). The precise nature of elements controlling transcription of these inflammatory mediators is for the most part unknown. It is almost certain, however, that their expression is primarily controlled by transcription factors that act by binding to the promoter regions of inflammatory genes. Nuclear factor- κ B (NF- κ B), an inducible nuclear *trans*-acting factor

known to play a key role in regulating expression of many inflammatory mediators, is the most strongly implicated of such transcription factors. Hyperactivity of NF- κ B has been implicated in a wide variety of chronic inflammatory diseases including asthma, rheumatoid arthritis and IBD (reviewed by Barnes and Karin (46)).

A wide variety of stimuli have been shown to activate NF- κ B, including cytokines, activators of protein kinase C, viruses, and oxidants (46). Some of the most well-characterized activation pathways appear to involve receptor-mediated events. Notable examples include signaling through the IL-1 and TNF α cytokine receptors, as well as microbial pattern recognition receptors, including Toll-like receptors and the intracytoplasmic bacterial peptidoglycan receptor, NOD2.

NF- κ B is highly up-regulated at sites of inflammation, and its activation induces transcription of pro-inflammatory cytokines, chemokines, adhesion molecules, matrix metalloproteinases, cyclo-oxygenase-2, and inducible nitric oxide synthase (47). The role of NF- κ B in IBD is well defined. Lamina propria macrophages display inappropriate activation of a number of NF- κ B subunits including NF- κ B1, c-Rel, and especially RelA (48). Treatment of a macrophage cell line derived from inflamed IBD mucosa with a specific NF- κ B p65 subunit antisense phosphorothioate oligonucleotide results in a significant reduction in pro-inflammatory cytokine production (48).

Innate immunity and a prominent component of this system that activates NF- κ B and other proinflammatory mediators, the Toll-like receptor (TLR) pathway will now be examined. TLR stimulation induces a strongly pro-inflammatory intracellular signaling pathway. The possible role that this pathway may play in IBD pathogenesis will then be assessed in depth.

Innate Immunity

Vertebrate host defense mechanisms can be broadly divided into two systems: ‘adaptive’ immunity and ‘innate’ immunity. Both serve to recognize and respond to invading pathogens by initiating host defenses. How each of these systems react and respond to different pathogenic threats are markedly dissimilar.

Adaptive immune recognition relies on generating a random and highly diverse repertoire of antigen receptors, the T- and B-cell receptors. This is followed by clonal expansion of receptors with pathogen-specific recognition abilities (49). Although this is an extremely powerful immunological defense mechanism, it has a number of prominent limitations, the most notable of which is that responses are slow. The lag-time for an effective response to a pathogen through adaptive immune mechanisms is typically 4-7 days. The primary reason for this slow response time is that activation of adaptive immunity requires signals from a variety of sources. Such signals provide information regarding the nature of the pathogenic threat, and serve to direct the type of response mounted to any given pathogen. These signals primarily originate from the innate immune system (50).

The innate immune system comprises an evolutionarily ancient host defense mechanism that shows a great deal of molecular conservation between plants to mammals. Innate and adaptive immunity differ in two broad respects. First, innate immunity comprises a system of effector mechanisms that, when activated, often lead to the swift destruction of an invading organism (typically within a matter of hours). Second, rather than mounting responses to a specific antigen, innate immune mechanisms respond to structures present on large groups of microorganisms and are distinct from

molecular patterns found in vertebrates (51). The targets of innate immune recognition are highly conserved molecular structures known as pathogen-associated molecular patterns (PAMPs) (49). The principal cells involved in initiating and mediating these responses are phagocytes such as neutrophils and macrophages, as well as a wide variety of other cells derived from non-leukocytic tissue. Once the phagocytic cell has reached the site of infection, microbial PAMPs are recognized by innate immune system receptors, so-called pattern-recognition receptors (PRRs) (52). The phagocyte then sets in motion a number of mechanisms that ultimately lead to the destruction of the pathogen and/or activation of the adaptive immune system.

Toll-Like Receptors

The *Drosophila* Toll protein and its mammalian homologues, the Toll-like receptors (TLRs), are evolutionarily conserved forms of PRR, and are essential components the innate immune system. They are expressed on a broad range of immune and non-immune system effector cells, and possess the ability to recognize a wide variety of microbial PAMPs. TLR stimulation activates a number of interlinked intracellular signaling pathways, all of which lead to innate and ultimately adaptive immune system activation (53).

Toll-like receptors are a family of ten structurally related type-1 transmembrane proteins (TLR1-10) (49). The TLR extracellular region contains multiple leucine-rich repeats (LRRs), a ligand recognition and signal transduction motif found in a diverse range of proteins (54). The LRR domains consist of a varying number of repeats of 24-29

amino acids in length, containing the motif xxLxLxx and other conserved leucines (55). The cytosolic region of TLRs is called the Toll/interleukin-1 receptor (TIR) motif owing to the structural homology of a 200 amino acid conserved domain found in both TLRs and interleukin-1 receptors (IL-1R) (56). TIR motifs are conserved protein-protein interaction domains that are found in both plant and animal transmembrane and cytosolic proteins (57). TIR domain-containing proteins have been shown to have a role in host defense in both plants and animals.

Toll-like receptors form a powerful means for innate immune effector cells to differentiate foreign versus host molecular patterns. More specifically, TLRs recognize a broad range of conserved molecular motifs found in microorganisms but not in vertebrates (56). The range of ligands recognized by TLRs is very diverse. For example, TLR2, TLR3, TLR4, TLR5, TLR6 and TLR9 recognize peptidoglycan, double stranded RNA, lipopolysaccharide (LPS), flagellin, mycoplasmal macrophage-activating lipopeptide-2kDa (MALP-2) and CpG bacterial DNA respectively (58-63). Ligand specificities for a variety of TLRs are shown in Figure 1. The main strategies for increasing the range of ligands recognized include formation of heterodimers (e.g. between TLR2 and TLR6) and the use of accessory molecules (e.g. TLR4 requires the accessory molecule MD-2 for recognition of gram negative bacterial LPS).

Tissue expression patterns of TLRs vary, with most tissues expressing at least one TLR and several expressing all ten (e.g. spleen, peripheral blood leukocytes) (64). Analysis of TLR expression in fractionated primary human leukocytes (CD4⁺, CD8⁺, CD19⁺ monocytes and granulocytes) indicates that professional phagocytes express the greatest variety of TLRs. However, expression of several TLRs appears more restricted to

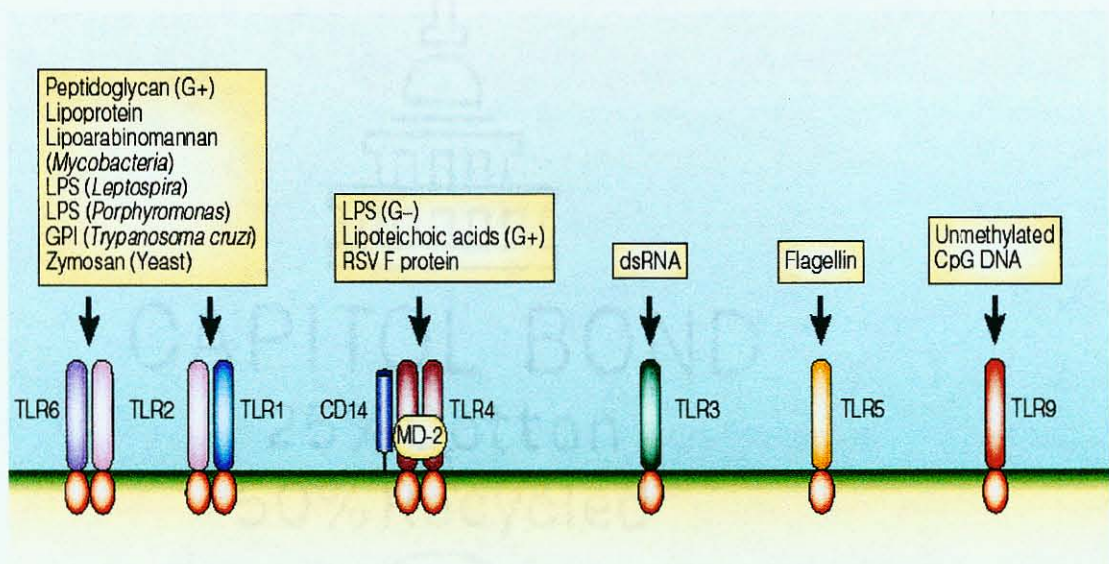


Figure 1. Ligand specificities for different TLRs. TLRs recognize a variety of PAMPs *in vivo* by forming homo- and heterodimers. Note that TLR2 forms a heterodimer with either TLR1 or TLR6. Such strategies increase the number of recognized PAMPs by allowing for more combinations of TLRs than would be possible if TLRs acted in isolation or only formed homodimers. Note that some TLRs can utilize accessory molecules (e.g. MD-2 with TLR4). The precise role of such accessory proteins is unclear in many instances. G+, gram-positive; G-, gram-negative; GPI, glycosphosphoinositol; RSV, respiratory syncytial virus. From Medzhitov R, 2001. Toll-like receptors and innate immunity: Nat.Rev.Immunol., v. 1, p. 135-145.

B cells, suggesting additional, and as yet poorly defined, roles for TLRs in adaptive immunity.

TLR Family Signaling Pathways

TLR stimulation leads to up-regulation of a number of cytokines involved in processes such as inflammation and chemotaxis. Much work has focused on defining specific intracellular signaling pathways that mediate such innate immune system effector cell responses. As with many other cell signaling pathways, a great deal of complexity is evident, with multiple connections to different cell signaling components as well as ‘cross-talk’ with other receptor-mediated pathways.

Characterization of the TLR signaling pathway leading to up-regulation of NF- κ B owes a great deal to the discovery of structural homology between the intracellular domains of the IL-1R and TLRs (53). The IL-1R pathway was reasonably well characterized when Medzhitov et al (50) described the human homolog of the *Drosophila* Toll protein (65;66). Indeed, the primary signaling pathways induced by activation of either IL-1R or TLRs appear to have the same end-points. These well-described signaling pathways are shown in Figure 2.

Both IL-1R and TLR stimulation lead to the recruitment of MyD88, a cytoplasmic adaptor protein that can interact with every member of the TLR family, as well as the IL-1R (65;67-69). MyD88 is essentially a C-terminal TIR motif, connected to another structurally conserved signaling motif, a death-domain (DD), which is located at its N-terminal (68). The TIR domain of MyD88 enables it to interact with the complementary

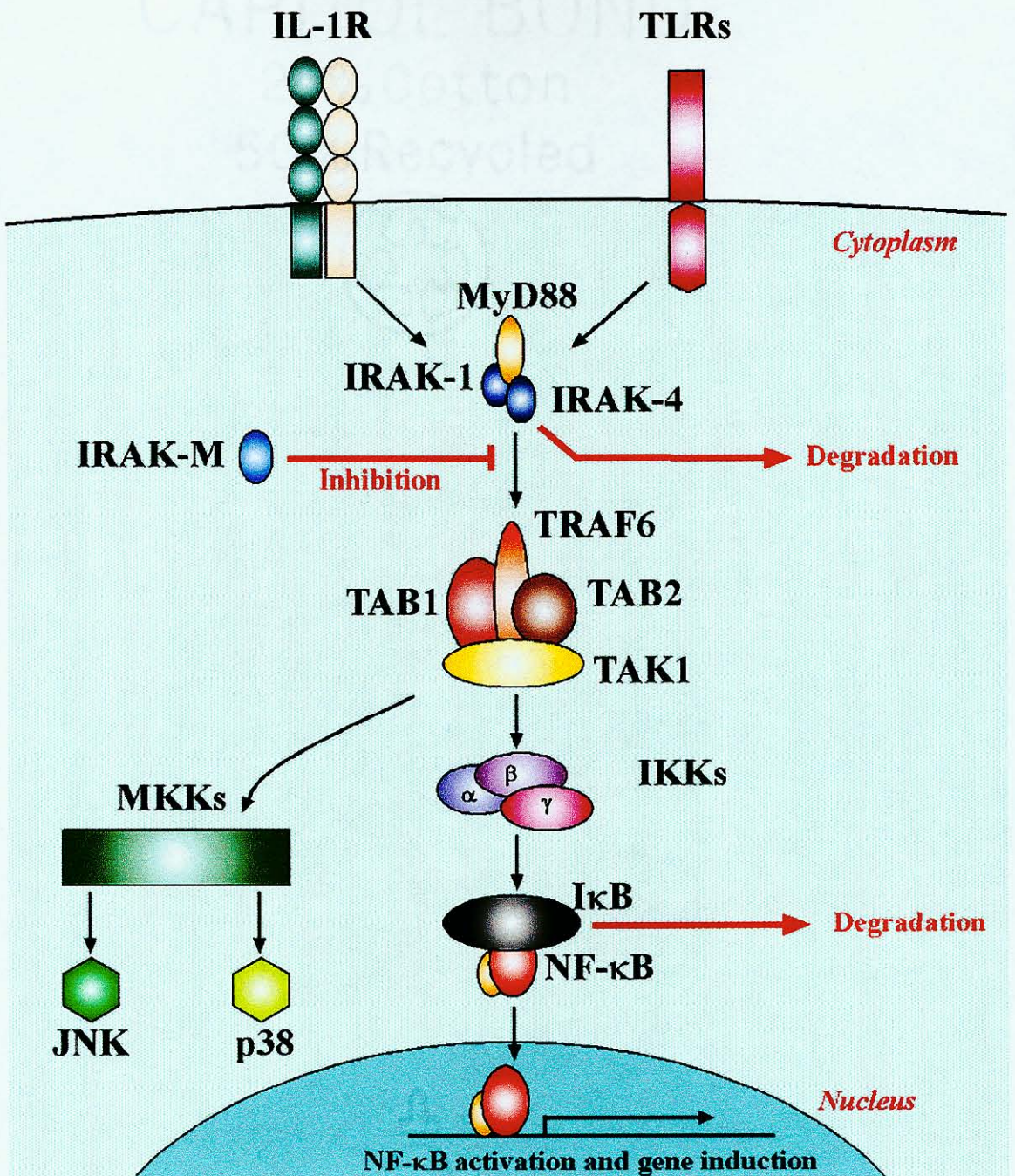


Figure 2. The interleukin-1 receptor (IL-1R)/Toll-like receptor (TLR) signaling pathway.

Adapted from Akira S. and Sato S., 2003. Toll-like Receptors and Their Signaling Mechanisms, Scand. J. Infect. Dis. 35:555-562.

TIR domain of ligand-bound TLRs (or IL-1R). Once recruited to the receptor complex, MyD88 binds the serine/threonine kinases IL-1R associated kinase-1 (IRAK-1) and IRAK-4 through DD-DD interactions. This subsequently causes IRAK-4 to hyperphosphorylate IRAK-1, which induces interaction with another adaptor protein, tumor necrosis factor receptor-associated factor 6 (TRAF6) (55). The association of these molecules presumably causes conformational changes, and the signaling module IRAK-4/IRAK-1/TRAF-6 dissociates from TLR/MyD88 (55). The downstream targets for this three-molecule complex include mitogen-activated protein kinase kinase kinase (MAP3K), transforming growth factor- β -activated kinase-1 (TAK1), and two adaptor proteins, TAK1-associated binding proteins-1 and -2 (TAB1 & 2).

Thus IRAK-4/IRAK-1/TRAF-6 interacts with membrane bound TAB2, and then the TAB2/TAK1/TAB1/TRAF6 complex translocates to the cytosol, leaving the IRAKs to be degraded. TAK1 is subsequently activated, which in turn leads to activation of yet another protein kinase, I κ B kinase (IKK) (51). IKK phosphorylates the Inhibitor of κ B (I κ B) proteins, which are bound to NF- κ B and inhibit its action in the un-stimulated state. Phosphorylation of I κ B targets it for ubiquitination and proteosomal degradation allowing NF- κ B to translocate into the nucleus where it induces the transcription of inflammatory and immune response genes. Activation of TAK1 also results in activation of MAP kinases.

This pathway is by no means the only signaling transduction cascade that may be induced through TLR stimulation. Much attention has recently focused on those TLR signaling pathways that are independent of the adaptor protein MyD88. This has stemmed from observations that MyD88 knockout mice are still able to activate NF- κ B

and MAP kinases in response to LPS, albeit in a delayed fashion (70). MyD88-independent pathways have been shown to activate interferon-inducible genes and promote LPS-mediated maturation of dendritic cells (56). These pathways do, however, remain of secondary importance to the MyD88-dependant pathway that is essential for induction of inflammatory responses to LPS.

Another interesting phenomenon elicited by macrophage TLR stimulation is the induction of apoptosis. Specifically, it has been observed that bacterial lipoprotein-stimulated TLR2 can either induce apoptosis or activate NF- κ B in macrophage cell lines (71). TLR signaling pathways leading to either inflammation or apoptosis appear to bifurcate at the level of MyD88. TLR2 has been shown to bind Fas-associated death domain protein (FADD) in a MyD88-dependant fashion (72). FADD is a universal apoptotic adapter protein that mediates signaling between all known death-domain-containing members of the TNF receptor super-family. The implications of TLR-mediated apoptosis are especially intriguing. One possible explanation is that activation of programmed cell-death may serve to limit inflammatory responses. That is, to resolve inflammation, immune effector cells must be eliminated through apoptosis (73).

The function of a number of TLR signaling pathway components will now be considered. Discussion will concentrate on two IBD candidate genes within this pathway, *IRAK-M* and *MyD88*. The possible roles of these genes in the pathogenesis of IBD will be discussed in Chapter III.

Interleukin-1 Receptor-Associated Kinases

Interleukin-1 receptor associated kinases (IRAKs) are a family of serine/threonine kinases that are integral components of the IL-1R and TLR signaling pathways. The first member of this family to be described was IRAK-1, which shares a significant deal of homology with the *Drosophila* protein Pelle, a kinase that is essential for activating the *Drosophila* NF- κ B homolog Dorsal (74). In addition to IRAK-1, three further IRAKs have been identified in vertebrates: IRAK-2, IRAK-4 and IRAK-M (69;74-76). All IRAKs share two structural domains: a centrally located serine/threonine-rich kinase domain and an N-terminal death-domain (DD).

Activation of either the TLR or IL-1R pathway induces the recruitment of IRAK-1 to the receptor through a homophilic interaction with the DD of the adaptor protein MyD88 (56). MyD88 also binds IRAK-4, which is brought into close proximity to IRAK-1. IRAK-4 is thus allowed to phosphorylate critical residue(s) in the kinase activation loop of IRAK-1, triggering the kinase activity of IRAK-1. This causes IRAK-1 to autophosphorylate residues at its N-terminus (56), which leads to the recruitment of the adaptor protein TRAF6 to the receptor complex.

Early studies implied that there might be a degree of functional redundancy between individual IRAK family members. This stemmed from the observation that all IRAKs act downstream of MyD88 and can bind to TRAF6. However, although different IRAK isoforms share a degree of structural homology, functional differences are apparent between individual family members. One of the most important differences is their varying kinase activity. IRAK-4 has been shown to have a similar level of kinase activity as IRAK-1 (75). However, IRAK-2 and IRAK-M both lack the kinase activity of IRAK-1

and IRAK-4. Furthermore, their ability to become activated through auto-phosphorylation is practically zero, and they only possess a partial ability to stimulate NF- κ B when over-expressed in 293T cells, and restore IL-1 signaling in IRAK^{-/-} cells (69;76). IRAK-2 and IRAK-M presumably lack kinase activity because they have an asparagine and serine residue, respectively, in their kinase domains in place of an aspartate residue shown to be critical for the kinase activity of other IRAKs (56).

It has been postulated that IRAK-4 is the central TIR-signaling mediator in innate immunity (77). This has stemmed from a number of observations including that IL-1R and TLR signaling are inhibited by expression of an IRAK-4 kinase-inactive mutant, with IRAK-4 deficient animals displaying severely impaired responses to viral and bacterial challenges. The same cannot be said of IRAK-1, where gene-targeting studies have shown that it is only partially required for IL-1R and TLR signaling (75;78). Similar kinase domain point mutations of other IRAKs do not lead to a similar severe degree of impairment of NF- κ B activation as that which occurs following point-mutation of IRAK-4. Suzuki et al (78) showed that IRAK-4 is necessary to functionally modify IRAK-1 and other signal transducing substrates. In addition to these observations, tissue expression patterns of the IRAKs differ with IRAK-1 and IRAK-4 being expressed ubiquitously (69;75), IRAK-2 having a narrower cellular distribution, and expression of IRAK-M being limited to myeloid cells (76).

The mechanisms underlying regulation of IRAK function are only just becoming apparent. IRAK-1 is known to be quickly degraded following TLR/IL-1R stimulation (79;80). It is likely that IRAK-1 is degraded through a phosphorylation-induced ubiquitin/proteosome pathway (80). Hu et al (81) demonstrated that LPS-induced

degradation of IRAK-1 requires a 186 amino acid residue region located at the N-terminal. Conversely, IRAK-2 was not degraded when THP-1 cells were exposed to LPS, suggesting complex regulation of innate immunity upon microbial stimulation.

IRAK-M: An Inhibitor of TLR Signaling

Kobayashi et al (58) described the apparent function of IRAK-M in a mouse strain that expressed functionally inactive IRAK-M. A surprising result was obtained upon TLR stimulation of macrophages derived from these mice, with IRAK-M^{-/-} macrophages displaying increased cytokine production and inflammatory responses.

Two approaches were used to determine the function of IRAK-M. First, murine IRAK-M^{-/-} macrophages were stimulated with either live or heat-killed bacteria. These cells displayed increased pro-inflammatory cytokine production in comparison to wildtype macrophages. Second, *in vivo* studies utilized IRAK-M^{-/-} mice infected with the enteric pathogen, *Salmonella typhimurium*. Such mice exhibited an exaggerated gastrointestinal inflammatory response, which again, is likely due to enhanced TLR signaling. These studies concluded that IRAK-M is a negative regulator of TLR signaling and is thus a key component in the regulation of the innate immune response.

Another interesting concept to consider in light of this work is the phenomenon of lipopolysaccharide (LPS) tolerance. When an individual is repeatedly exposed to LPS, the threshold of LPS response is increased (82), representing a form of negative feedback that presumably has evolved to protect the organism against excessive cytokine production, and the potentially fatal endotoxic shock (83). The exact mechanisms

involved in this phenomenon remain uncertain. However, several factors have been proven to be involved in this process. For instance, TLR4 is down-regulated following chronic LPS exposure (84), as well as repression of transcription and rapid degradation of pro-inflammatory cytokine mRNAs in LPS-tolerant cells (85). The elucidation of the function IRAK-M by Kobayashi et al (58) has added another dimension to the mechanism of endotoxin tolerance. Kobayashi et al (58) found that IRAK-M^{-/-} macrophages have diminished ability to develop LPS tolerance after repeated exposure to LPS. There was also increased expression of IRAK-M following exposure of wildtype macrophages to high levels of LPS. IRAK-M^{-/-} macrophages did retain some capacity to develop LPS tolerance at late time points. It was thus concluded that IRAK-M up-regulation in response to high LPS loads is partially responsible for inducing endotoxin tolerance.

The gene encoding IRAK-M was isolated from a full-length cDNA clone by Wesche et al (65). The protein itself is 596aa in length and has 30-40% overall sequence similarity with IRAK-1 and IRAK-2. *IRAK-M* is located on chromosome 12q13.13 and is in extremely close proximity to the peak region of linkage within the IBD2 susceptibility locus. The gene itself consists of 12 exons and spans approximately 60kbp.

Mechanism of Action of IRAK-M

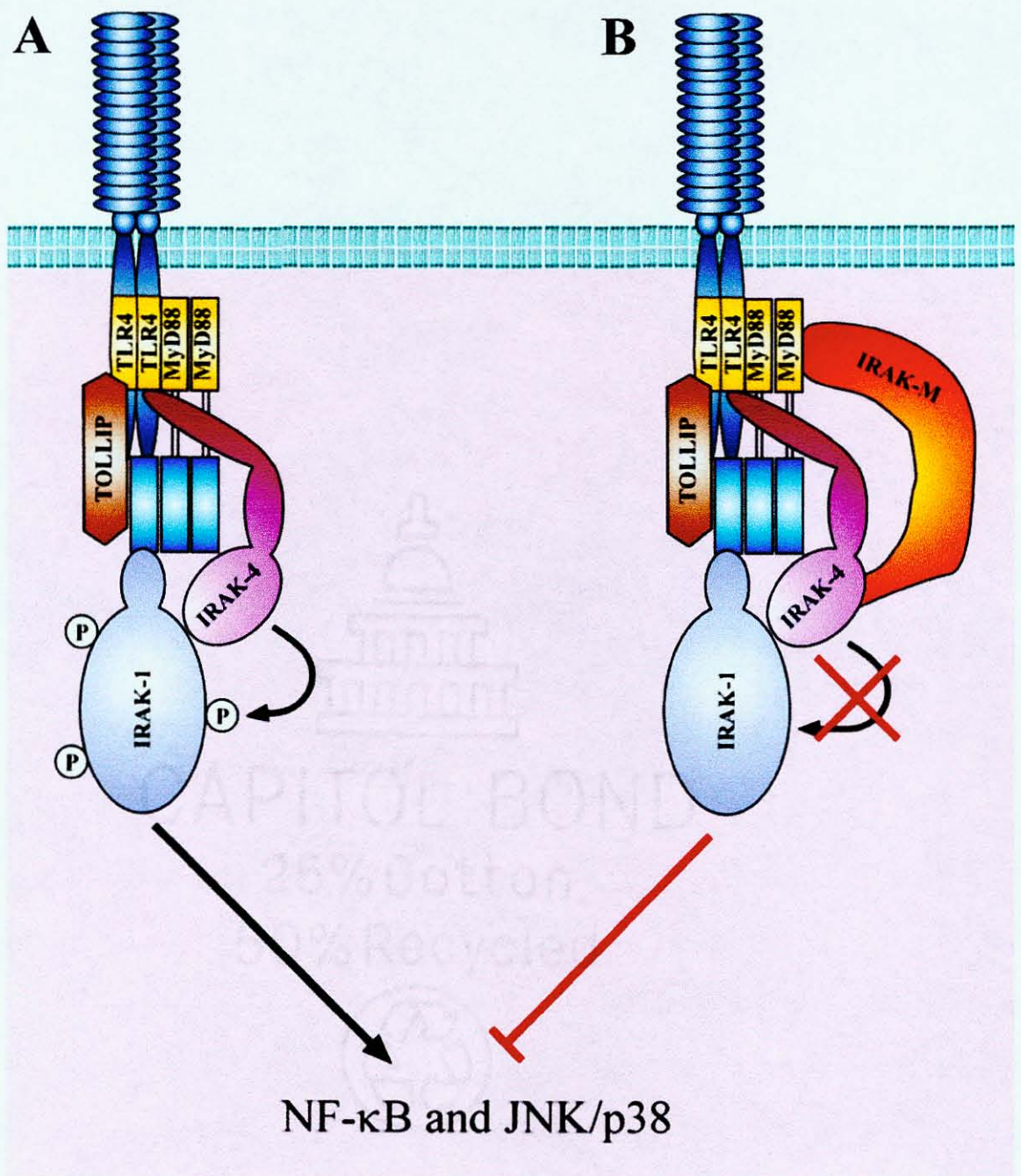
Kobayashi et al (58) proposed a mechanism of action for IRAK-M, and although by no means biologically implausible, it is somewhat a matter of speculation. The proposed mechanism of action of IRAK-M is shown in Figure 3. The action of IRAK-M

Figure 3. Proposed mechanism of action of IRAK-M

A: Uninhibited TLR4 signaling. Following stimulation by a PAMP, MyD88 binds the TLR (TLR4 in this instance) through a TIR interaction, and recruits kinases IRAK-1 and IRAK-4 to form an active signaling complex. The proximity of these two kinases causes hyper-phosphorylation and a subsequent conformational change in IRAK-1. IRAK-1/-4 binds the adaptor protein TRAF-6, which causes IRAK-1/IRAK-4/TRAF-6 to be released from TLR4/MyD88 and to interact with downstream molecules. Note that IRAK-1 is bound to TOLLIP. The function of this molecule is unknown, although it seems likely that it is either an inhibitor of unphosphorylated IRAK-1 or another adaptor protein.

B: Inhibition of IRAK-1 by IRAK-M. In the presence of IRAK-M, TLR stimulation again leads to the formation of the TLR-MyD88-IRAK-1/-4 complex. However, on this occasion IRAK-M also joins the complex. This has two proposed effects: it either stabilizes the signaling complex, preventing dissociation of IRAK-1 from MyD88, or inhibits phosphorylation of IRAK-1 by IRAK-4. Thus IRAK-1 is not freed to interact with downstream signaling intermediaries.

Adapted from Janssens S. and Beyaert R., 2003. Functional Diversity of Different Interleukin-1 Receptor Associated Kinase (IRAK) Family Members. *Molecular Cell* 11:293-302.



was proposed on the basis of three observations: 1) the presence of IRAK-M enhanced the interaction between IRAK-1 and MyD88; 2) the auto-phosphorylative capacity of IRAK-1 was significantly diminished by IRAK-M; and 3) IRAK-M could inhibit the interaction of IRAK-1 with its downstream target TRAF6. The proposed primary mode of action of IRAK-M was by binding to the TLR-MyD88-IRAK-1/-4 signaling module. This would have dual effects: to stabilize the signaling complex, preventing IRAK-1/-4 from being released to act on its downstream substrates; and to inhibit the auto- or cross-phosphorylation of IRAK-1/-4, thus preventing activation. Additionally, IRAK-M has previously been shown to be capable of binding TRAF6 (76). It is plausible that IRAK-M may compete with IRAK-1/-4 by binding to this adaptor protein.

However, the example shown above is certainly not the definitive explanation for the mechanism of action of IRAK-M. Mak and Yeh (86) implied that the findings of Kobayashi et al (58) raise a number of other important questions. Pro-inflammatory signaling pathway activation following IL-1R/TLR stimulation is limited by the rapid degradation of phosphorylated IRAK-1 (79), thus quickly depleting the intracellular pool of IRAK-1. Consideration of Kobayashi's data leads to the question of whether phosphorylation of IRAK-1, its activation status and its ability to be degraded are altered by the presence of IRAK-M. These questions remain unanswered.

MyD88 and TLR-Mediated Apoptosis

MyD88 is an essential component of a number of receptor-mediated cascades including the IL-1R and TLR pathways. Burns et al (67) showed that IL-1-induced NF-

κ B activation was inhibited by transfecting cells with a dominant negative mutant form of MyD88. Furthermore, they illustrated that over-expression of MyD88 led to NF- κ B activation. The role of MyD88 in TLR signaling was initially described by Medzhitov et al (68), where it was found to mediate activation of NF- κ B via IRAK-1 and TRAF6. Further evidence was provided by Kawai et al (70), who demonstrated that MyD88-deficient mice are unresponsive to LPS.

MyD88 is a 296-amino acid polypeptide with a predicted mass of 33 kD (87) consisting of a carboxyl-terminal TIR motif, connected to an amino-terminal death-domain (DD) (68). The TIR domain of MyD88 enables it to interact with the complementary TIR domain of ligand-bound TLRs (or IL-1R). Once ligand-activated TLR has bound MyD88, the serine/threonine kinases IRAK-1 and IRAK-4 are recruited through a DD-DD interaction with MyD88. The *MyD88* gene consists of 5 exons and is located on chromosome 3p22-3p21.3. This is a noteworthy point; chromosome 3p has been previously linked to IBD susceptibility (IBD9 susceptibility locus) (19;22;88).

MyD88 is not only involved in NF- κ B activation following TLR and/or IL-1R stimulation. It has also been shown to be an intermediary in a pathway leading to macrophage apoptosis. This was concluded from experiments that involved exposing TLR2-expressing macrophages to the TLR2 ligand, bacterial lipoprotein. These cells would either produce inflammatory responses via the familiar NF- κ B pathway, or would undergo apoptosis (71). TLR2 signaling appears to bifurcate at the level of MyD88. Additionally, TLR2 has been shown to bind Fas-associated death domain protein (FADD) in a MyD88-dependant fashion (72). FADD is a universal apoptotic adapter protein that mediates signaling between all known death domain-containing members of

the TNF receptor superfamily. Inhibition of the inflammatory pathway appears to enhance activation of the apoptotic pathway and vice versa. Regulation of the balance between these differing pathways *in vivo* is, however, poorly understood.

The implications of TLR-mediated apoptosis are especially intriguing. One possible explanation is that the activation of programmed cell-death may serve to limit the inflammatory response. That is, to resolve inflammation, immune effector cells must be eliminated through apoptosis (73). The role of MyD88 in apoptotic and inflammatory TLR-activated pathways is shown in Figure 4.

Further evidence for the existence of apoptotic pathways reliant on MyD88 have been provided by a number of studies. The phenomenon of TLR-mediated macrophage apoptosis (71) was verified by Ruckdeschel et al (89) who investigated the mechanisms by which the *Yersinia enterocolitica* outer protein P (YopP) interferes with the host immune response. They demonstrated that not only did YopP inhibit activation of NF- κ B in macrophages, but also induced a pro-apoptotic signal relay through MyD88 and IRAK-2, which potentially targets the FADD/caspase-8 apoptotic pathway. Conversely, IRAK-1 and TRAF6 were shown to counteract the bacterially induced cytotoxic response by signaling macrophage survival.

In addition to these studies, ectopic expression of MyD88 in human cervical carcinoma HeLa cells induced apoptosis, thus further emphasizing the unusual role that this protein assumes in transduction of programmed cell death signaling pathways (90;91). Recently, MyD88 was also documented to mediate signaling in Taxol-induced apoptosis associated with TNF α expression in human myelomonocytic cells (91;92).

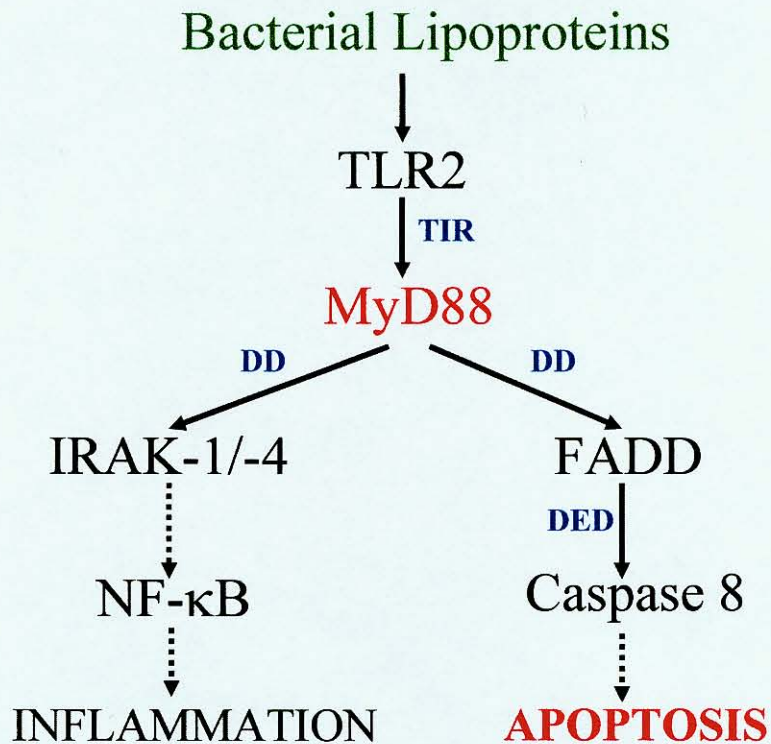


Figure 4. TLR-activated inflammatory and apoptotic pathways bifurcate at the level of MyD88 in macrophages. The death domain (DD) of MyD88 most often interacts with the DD of IRAK-1/-4, thus causing NF-κB up-regulation and inflammation. Alternatively, in certain circumstances, MyD88 can interact with the DD of Fas-associated death domain protein (FADD). FADD in turn interacts with caspase 8 through its death effector domain (DED) motif. Caspase activation initiates apoptotic signaling cascades within the macrophage. **Solid line** = Direct protein-protein interaction. **Broken line** = Indirect interaction.

In conclusion, MyD88 appears to be involved in induction of apoptosis in macrophages as a consequence of TLR stimulation, as well as playing a significant role in inflammatory responses through activation of transcription factors such as NF- κ B. This TLR-induced apoptosis is of somewhat unclear biological significance, but is likely involved in resolution of the inflammatory response. It is therefore quite evident that disruption of this process may have drastic consequences, especially in the context of inflammatory diseases.

CHAPTER III

FUNDAMENTAL QUESTIONS

IBD develops in a genetically susceptible individual in response to environmental stimuli. This type of inheritance pattern is described as ‘complex’ or ‘non-Mendelian’. That is, a genetically susceptible individual will carry multiple abnormal genes, which in isolation may have little or no phenotypic effect. However, it is the combination of these variant genes, and how they interact that determines phenotype.

Progress in defining IBD susceptibility genes has proven extremely difficult. This situation is not unique to IBD and can generally be considered to be the case for all complex diseases. Despite elucidation of multiple IBD susceptibility loci, only one IBD susceptibility gene, *NOD2*, has thus far been identified (18-25). *NOD2* is a component of the innate immune system, the body’s first line of defense against invading pathogens. Mutations within *NOD2* increase susceptibility to a number of distinct CD phenotypes in Caucasian populations (26;27;31-36).

This dissertation primarily focuses on defining new IBD susceptibility genes. The candidate genes to be studied are components of the Toll-like receptor pathway, an integral component of the innate immune system. The innate immune system has the capacity to induce powerful inflammatory responses when provoked, and there is accumulating evidence to suggest that innate immune dysfunction may be a key initiating factor in IBD pathogenesis. TLR stimulation activates a strongly proinflammatory

intracellular signaling pathway that leads to up-regulation of NF- κ B, a nuclear *trans*-acting factor induced upon stimulation of the innate immune system. NF- κ B is known to play a key role in regulating expression of many inflammatory mediators, and hyperactivity of this transcription factor has been implicated in a wide variety of chronic inflammatory diseases including asthma, rheumatoid arthritis and IBD (46). Additionally, in certain circumstances, TLR stimulation cause programmed cell death in macrophages, presumably as a measure initiated to limit the duration and intensity of inflammatory responses.

Defining susceptibility genes in IBD, as in all complex diseases, is of paramount importance. Elucidation of such genes will improve our understanding of pathophysiologic mechanisms underlying disease development, and may enable the development of novel therapeutic interventions or more efficient means of diagnosis.

Proposed Concepts

This dissertation addresses six proposed concepts: 1) Not all of the nine or more previously described IBD susceptibility loci will be associated with disease susceptibility in any given population. This is primarily due to intrinsic population differences. 2) The high consistency of reliable phenotyping in the study population will allow for more accurate assessment of disease associations. Previous IBD susceptibility loci linkage studies have produced variable results either because of inconsistent phenotyping of disease populations, or because IBD populations have been considered as large, heterogeneous groups (i.e. not considering sub-phenotypes). 3) IBD susceptibility loci are

associated with disease because they contain susceptibility genes. 4) TLR signaling pathway dysfunction is an integral component of IBD pathogenesis since inappropriate activation of these receptors may have wide-ranging effects, including hyper-stimulation of inflammatory and apoptotic signaling cascades. 5) *IRAK-M*, a gene that lies within a previously described IBD susceptibility locus whose product is a potent inhibitor of TLR signaling, is a biologically plausible IBD candidate gene. 6) Given the importance of the TLR pathway in innate immune system homeostasis, other components of the intracellular signaling are system biologically plausible candidates, the most notable of which is the adaptor protein MyD88.

To address these six concepts, a series of experiments were designed to test the following hypotheses.

Hypothesis #1

A variable number of IBD susceptibility loci will be associated with disease in this population. That is, due to intrinsic population differences and high consistency of phenotyping, not all of the nine or more loci described in different populations will have an apparent role in disease susceptibility here.

Some susceptibility loci have been described in very small, isolated populations. For example, the strongest evidence for linkage of the IBD7 locus on chromosome 1p36 has been described in an American Chaldean population, a Roman Catholic group from Iraq (93). Other susceptibility loci have been described in broad, somewhat different populations than is the case in this study. Examples of such loci include the IBD9 locus

on chromosome 3p which was described in a Northern European Caucasian population derived from many countries (88). Furthermore, some loci have only been described in one study, and have never been replicated. For example, the locus on chromosome 4q has only been described once, in a North American Caucasian population (23).

Such inconsistent findings from linkage studies have proved a prominent confounding factor in the search for susceptibility genes. To test this hypothesis, experiments will be designed to define disease associations for a number of susceptibility loci in the patient population. This will be achieved by characterizing peak LOD score STR markers from previous studies (i.e. those STRs most highly linked to disease). Genotype and allele frequencies will be compared between disease and control populations through the use of population- and family-based statistical approaches.

This approach will have distinct advantages over these linkage studies. Firstly, individuals within this IBD population generally have a severe affection status, and are from the relatively small geographic region of Kentucky and Southern Indiana. The study population is therefore more clinically homogenous when compared to other cohorts. Secondly, disease groups will be analyzed at the sub-phenotype level. Experience with *NOD2* has shown that higher degrees of association are achieved using this approach, than when considering unstratified IBD groups. Finally, both population- and family-based methodology will be employed to analyze STR data. These approaches are complementary and can be considered different means of approaching the same problem. Family-based approaches account for population substructure, a significant confounding variable that can cause type 1 statistical error in population-based studies.

Hypothesis #2

The IBD2 susceptibility locus is associated with disease in the Kentucky and Southern Indiana IBD patient population. Furthermore, this disease association is due to the presence of *IRAK-M* within IBD2.

Previous studies of IBD2 in this patient population have demonstrated that IBD2 appears to be associated with UC (94). The aim here is essentially to extend the previous study by performing a more detailed characterization of the IBD2 locus, thereby determining the role it plays in IBD susceptibility. Three STR markers will be characterized in the same severely affected population in distinct phenotypic groups. The experimental procedure and the approach to analysis of STR data will essentially be the same as used in Hypothesis #1.

The IBD candidate gene *IRAK-M* lies within 5Mb of the peak region of linkage of IBD2 (both are located on chromosome 12q13). The finding of IBD2 disease association will be seen as providing evidence to suggest that *IRAK-M* is a prudent choice as an IBD candidate gene. Detailed characterization of IBD2 is therefore considered as an essential preliminary step prior to assessing *IRAK-M* an IBD candidate gene.

Hypothesis #3

Mutation of *IRAK-M*, a potent inhibitor of TLR signaling, lessens its ability to negatively regulate this signaling pathway in IBD. Bowel luminal contents therefore over-stimulate TLRs present on bowel lamina propria macrophages, which in turn produce excessive amounts of proinflammatory mediators through

activation of NF- κ B and other transcription factors. Prolonged over-production of proinflammatory cytokines, chemokines and other inflammatory mediators in the gastrointestinal tract will eventually cause the type of chronic inflammation characteristic of IBD.

IRAK-M is a biologically plausible IBD candidate gene. When considering the work of Kobayashi et al (58) that demonstrated the apparent importance of IRAK-M in regulation of innate immune responses (see p.24), it is obvious that even a mildly hypoactive form of this molecule may have drastic consequences. Activation of IRAK-M has a powerful anti-inflammatory effect through inhibition of TLR signaling. An individual in whom IRAK-M was hypoactive would be unable to fully ‘dampen’ TLR-mediated signaling. If an event occurred that necessitated the recruitment of gastrointestinal mucosal macrophages (e.g. mild injury or infection), exposure of these innate immune effector cells to normal bowel luminal contents would provoke an inflammatory response of an inappropriate magnitude. If this pattern were to be repeated over a prolonged period, it is logical to hypothesize that such an individual could develop the type of chronic inflammation characteristic of IBD.

A number of criticisms of this hypothesis do, however, need to be addressed. Firstly, given that IRAK-M appears to be important in developing LPS tolerance, why are IBD patients not more susceptible to developing endotoxic shock if IRAK-M dysfunction is indeed involved in IBD pathogenesis? Secondly, why is it that IBD patients do not mount inappropriate inflammatory responses to stimuli that cause TLR-activation in other organs? Both questions can be partially answered by considering the type of genetic variation found in complex diseases. It is unlikely that a mutated gene would be

completely non-functional in a complex disorder such as IBD (this situation is more likely to be the case in Mendelian disorders). Genes involved in complex disorders such as IBD are more likely to display more subtle functional abnormalities. Such a gene product may fulfill its normal biological role, but at a different level of functionality. In LPS tolerance, IRAK-M hypoactivity may be adequately compensated for by other mechanisms and LPS hyper-responsiveness would not be a feature of IBD. When considering the absence of inappropriate inflammatory responses in other organ systems in IBD, it should be remembered that the human gastrointestinal tract is exposed to chronically high microbial levels. If *IRAK-M* mutation were indeed involved in the IBD pathogenic process, the alimentary tract would be the most likely site of inappropriate inflammatory responses owing to the persistently high bacterial load present in the bowel lumen. IRAK-M under-activity in this type of environment could lead to chronic inflammation, whereas, other organ systems lacking similar levels of pathogen exposure would be apparently disease-free.

A two-fold approach will be used to test this hypothesis and to assess the feasibility of *IRAK-M* as an IBD candidate gene. Firstly, colonic mucosal expression levels of *IRAK-M* will be measured in both healthy and IBD-affected mucosa. Although abnormal gene expression levels in disease-affected mucosa are by no means categorical evidence of gene involvement in disease pathogenesis, it does imply that regulation of gene expression is abnormal, which in turn may be due to mutation of regulatory regions within that gene. This approach basically provides further evidence that the candidate gene is a prudent choice, and that more detailed characterization may be warranted. Analysis of gene expression will be performed using real-time PCR.

The second approach is to screen coding and regulatory genetic regions for single nucleotide polymorphisms (SNPs). SNPs located within such regions may be of great interest since they may cause abnormal gene function. Ideally, non-synonymous coding SNPs that cause changes in the amino-acid sequence of the encoded protein, or polymorphisms that may interfere with gene regulatory region functionality such as those within the promoter or 3'-untranslated region (3'-UTR) will be defined. Although of less interest, SNPs in intronic regions flanking exons will also be considered for further study since they may possess some form of regulatory role.

Once the screening process has been completed, identified polymorphisms will be defined in the entire study population. Statistical analysis methods similar to those used in the susceptibility loci studies will be employed to analyze polymorphism allele frequency data (i.e. population- and family-based methods).

Hypothesis #4

Mutation of MyD88 in IBD limits its ability to induce apoptosis upon TLR stimulation, thereby allowing inflammation to continue for an abnormal period of time.

As discussed above, MyD88 is not only involved in NF- κ B activation following TLR and/or IL-1R stimulation. Aliprantis et al (71) demonstrated that MyD88 is an intermediary in a pathway leading to macrophage apoptosis in response to TLR2 stimulation by its ligand, bacterial lipoprotein. The TLR signaling pathway appears to bifurcate at the level of MyD88. TLR2 stimulation has been demonstrated to activate the

proinflammatory transcription factor NF- κ B by binding IRAK-1 and IRAK-4. However, in certain circumstances MyD88 can bind Fas-associated death domain protein (FADD), a universal apoptotic adapter protein that mediates signaling between all known death domain-containing members of the TNF receptor superfamily (72). A number of plausible explanations for this intriguing finding have been suggested, the most likely of which is that activation of programmed cell-death in immune effector cells may serve to limit the inflammatory response. That is, to resolve inflammation, immune effector cells must be eliminated through apoptosis (73). It is easy to see that if this process were to be disrupted, immune effector cells would not be cleared following elimination of the pathogenic threat. Indeed, the hypothesis presented here states that this is the case in IBD. Immune effector cells within the bowel wall are not cleared in IBD since TLR-mediated apoptosis is defective. Inflammation will therefore continue for an abnormal length of time due to the prolonged stimulation and subsequent production of inflammatory mediators by immune effector cells.

It should be noted, however, that the case for *MyD88* as an IBD candidate gene does not appear to be as strong as for *IRAK-M*. This primarily relates to the genomic location of *MyD88*. The existence of an inflammatory bowel disease locus, IBD9, on chromosome 3p26 is supported by significant linkage, transmission/disequilibrium and partitioning of linkage evidence (25). *MyD88* is also located on the short arm of chromosome 3, although it is some distance from the peak region of linkage in IBD9, being located at 3p22. A significant genetic distance therefore lies between IBD9 and *MyD88*, and it is for this reason that *MyD88* appears the least likely candidate gene, despite appearing to be biologically plausible.

The approach taken to assess *MyD88* as an IBD candidate gene will be the same as for *IRAK-M*. The feasibility of *MyD88* as a candidate gene will be assessed by determining its colonic mucosal expression levels, and by screening coding and regulatory regions for SNPs. Following this, identified SNPs will be characterized in the entire population and SNP frequencies analyzed using population- and family-based statistical approaches.

Biological Plausibility and Candidate Gene Selection

To reiterate, biological plausibility is considered the most important factor in deciding which genes are to be the focus of this research. Indeed, biological plausibility should be assessed in all candidate gene studies by giving consideration to both the relevance of the proposed candidate in the pathogenesis of the particular disease, and to the functional effects (if any) of a polymorphism (95). The argument for the biological plausibility of genes included in this study is outlined above. Functionally significant polymorphisms rather than random polymorphisms in a gene of interest offer considerable advantages in terms of detecting disease-associated genes (95;96). It is for this reason that candidate genes will be screened for polymorphisms primarily within the coding and regulatory regions (i.e. promoter and 3'-UTR).

The major disadvantage of the candidate gene approach is that in any given complex disorder, numerous genes may appear to be biologically plausible candidates. It is, however, likely that the vast majority of such genes will not be disease-associated. Analyzing candidate expression levels in colonic mucosa of IBD patients will therefore

be used to narrow the choice of genes. If there is no significant difference in candidate gene expression (i.e. ≥ 2 -fold change in either direction in IBD-affected colonic mucosa when compared to normal colonic mucosa), that gene will be excluded from further study. Gene expression analyses will be performed for each biologically plausible candidate gene prior to more detailed characterization (i.e. screening for SNPs) using real-time PCR.

CHAPTER IV

MATERIALS AND METHODS

Patient Population and Study Participation

A population of IBD patients was recruited from a University surgical practice for study participation. These patients have been referred to a surgeon because their disease has not responded to medical therapy and they may therefore be classified as having severe disease. Patients originate from a relatively small geographical area, consisting of Kentucky and southern Indiana. These areas are predominantly rural, which is unusual since IBD is more prevalent in urban populations (13).

The initial diagnosis of IBD was determined through conventional radiological, endoscopic, and/or histopathologic findings. Histology was available in all cases. Following the initial diagnosis, a single gastrointestinal pathologist with a particular interest in IBD reviewed all histology. Previous studies have shown that interobserver variation in histological diagnosis can be a significant confounding variable (97). If there was disagreement between the initial diagnosis and that of the specialist pathologist, the latter diagnosis was used for purposes of this study. The approach used here was intended to maximize diagnostic consistency, thus increasing the clinical homogeneity of IBD subgroups.

Individuals without IBD also participated in this study. Information regarding race, ethnicity, and family history of IBD was obtained in a similar manner to IBD patients. These individuals will be derived from the same geographic area as the IBD patient population and will be recruited from patients without a family history of IBD who present for a screening colonoscopy. Since they are derived from the same practice, they should be similar to the IBD population with respect to geographic and socioeconomic variables. Due to the fact that they are presenting for cancer screening, many of these individuals will be older than the IBD patients. While this does represent a significant selection bias, it does have one important advantage. Specifically, since IBD may have a late age of onset, having controls from an older age group minimizes the chance that these patients will subsequently develop IBD, and therefore be misclassified as “unaffected”.

Patients with biopsy or surgically confirmed IBD, available histology, complete personal and family history, and signed written informed consent will be eligible for study participation. The following information will be documented for IBD patients: age, race, gender, age of diagnosis, family history of IBD and colon cancer, smoking history, number and type of previous operations, presence of extra-intestinal manifestations, and presence of colonic dysplasia or cancer. Disease behavior (uncomplicated, stenosing, or fistulizing disease) (98), presence and location of fistulae, presence or absence of perianal disease, and disease recurrence will also be assessed for CD patients. Such data will be of vital importance to this study since the overall aim is to define genotype-phenotype relationships for susceptibility loci and candidate genes.

IC was considered as a distinct IBD phenotype in this study. Strong evidence (99) has shown that IC should be considered a distinct clinical entity. IC patients have been shown to suffer from more fulminant disease and have a higher incidence of complications following surgery. One could argue that with sufficient follow-up, the majority of these patients will eventually be categorized as either CD or UC. This IC population has, however, been observed for a significant length of time, with an average 5.6 year follow-up and maximum follow-up of 12 years. IC patients exhibiting diagnostic features of either UC or CD are promptly re-categorized, thus increasing group homogeneity.

Eight milliliters of peripheral whole blood will be obtained for each study participant. Leukocyte DNA will be extracted (Puregene DNA isolation kit, Gentra Systems, Inc., Minneapolis, MN), quantified spectrophotometrically (GeneQuant, Pharmacia, Dublin, OH), and stored at 4°C until analysis. In patients undergoing surgery, 1cm² colonic mucosal samples will be harvested. Half the sample will be sent to pathology for hematoxylin and eosin staining and the other half immediately placed in RNLater (Quiagen, Valencia, CA) and stored at 4°C until analysis. In “control” patients who undergo hemorrhoidectomy, 1cm² “control” mucosal samples will be obtained and similarly processed.

Characterization of IBD Susceptibility Loci STR Markers

The primary goal of this part of the study is to perform case-control and family-based tests of association using STR markers within a number of IBD susceptibility loci.

Choice of STR markers for the preliminary IBD susceptibility loci study, and for the more detailed characterization of IBD2 will be discussed in Chapters V and VI respectively.

Polymerase chain reaction (PCR) primers will be custom synthesized (Proligo Corp, La Jolla, CA) and STRs amplified using a standard PCR protocol (Perkin-Elmer Thermocycler, Norwalk, CT) at amplicon-specific annealing temperatures and reagent concentrations. These conditions will be discussed in detail in the relevant chapter.

Amplicons will be analyzed using high performance liquid chromatography (HPLC [WAVE[®], Transgenomic, Omaha, NE]). Larger PCR fragments are eluted from the DNasep cartridge (WAVE[®], Transgenomic, Omaha NE) less rapidly than smaller alleles.

Statistical Analysis

Case-Control Tests of Association: Two-way contingency table analysis will be used to test for differences in allele and genotype frequencies among groups. Each marker will be tested separately for homogeneity of allele-genotype frequencies across groups by two-way contingency table analysis. Since there are many rare alleles and genotypes, alleles and genotypes with a frequency of less than 5% are pooled into one class for analysis.

The Mehta-Patel (100) extension of Fisher's exact test for two-way tables was used for computing P values. Case-control tests will be conducted in a three-step fashion. First, a Monte Carlo estimation of the global exact test for overall allele-genotype homogeneity will be calculated using 2 x c contingency tables for disease group versus

control populations (or disease *versus* disease in the case of CD disease behavior subgroups, and age category *versus* age category in the case of age of disease onset). Second, these P-values will be combined using Fisher's method (101) as a global test of allele-genotype homogeneity. Individual genotypes or alleles will only be tested when both the global and disease *versus* control comparisons achieve statistical significance following Bonferroni correction, thus accounting for the problem of multiple analyses and compounding of a type I error. Individual genotypes or alleles are compared using Fisher's exact test on a 2x2 table. Computations are performed using Statistical Analysis Software 9.0 (SAS Institute, Cary, NC).

Haplotype Estimations: Haplotype estimations will only be performed when appropriate i.e. when STR markers are on the same chromosome and are in close enough genetic proximity for linkage disequilibrium to exist between STRs. Haplotype frequencies for STR pairs will be estimated using an EM maximum likelihood method. The Markov test (Arlequin 2.0 software (102)) is used to test the differentiation between disease and control populations in haplotypic data. Maximum-likelihood estimates of haplotype frequencies will be calculated from genotypes of unknown phase, and standard deviations of estimates are obtained using a Monte Carlo method. The Markov chain will consist of 10,000 steps.

Family-Based Tests: Family-based tests of association are used to compare transmission of alleles from parents to affected siblings in different family units. They can also be used to compare allele transmission between affected and unaffected (phenotypically discordant) siblings. The Pedigree Disequilibrium Test (PDT) and a likelihood ratio test as implemented in the computer program TRANSMIT will be employed in this study.

The PDT allows for the use of genotype data from extended pedigrees, including triads and discordant sibling pairs, related nuclear families, and larger sibships from pedigrees. The PDT provides a valid test of both linkage and association (103). The PDT is a valid test of association, even if population substructure is present. Use of the PDT is expected to result in substantial gains in power. The PDT will remain robust, even when sample sizes as small as 25 pedigrees are used. The PDT-sum statistic, which will be utilized in this study, is a variation of the PDT used to correct for potential bias (104).

A likelihood ratio test as implemented in the computer program TRANSMIT will also be used as a method of family-based testing (105). It can deal with the transmission of multilocus haplotypes, even if the phase is unknown. If parental genotypes are unknown, data from unaffected or siblings of unknown status can be used to narrow the range of possible parental genotypes. For cases in which no information on parental genotype is available, the likelihood ratio test can be performed using data from either affected or non-affected siblings.

Apparent disease associations can occur in population-based studies due to either true disease association or spurious associations due to the effects of population substructure (106). Population substructure may occur if there are, for example, multiple population subtypes in what is assumed to be a homogeneous population. Family-based studies, unlike population-based approaches, have the advantage of not being affected by population substructure. Case-control studies are however superior in a number of ways, notably that they permit analysis of larger groups of patients than would typically be possible in linkage studies. Confirmation of results from the case-control population

study utilizing family-based testing will indicate if linkage and association are indeed present.

Characterization of TLR Pathway IBD Candidate Genes

Colonic Mucosal Expression of Candidate Genes

Candidate gene expression will be compared between colonic mucosa obtained from IBD patients and from subjects without IBD or colon cancer. These non-IBD, non-cancer mucosal samples will be obtained from individuals undergoing hemorrhoidectomy. In order to conclusively demonstrate a trend in gene expression, these experiments will be performed using tissue samples from several individuals with each type of IBD as well as in controls.

Following informed written consent, 1cm² of colonic mucosa will be obtained from IBD patients immediately following bowel resection, and placed in RNeasy[®] (Qiagen, Valencia, CA) and stored at 4°C. To ensure accurate disease assessment and uniformity of sampled tissue, matched tissue sections will be obtained for routine histological examination. Analyses will be performed only after a specialist gastrointestinal pathologist has verified the diagnosis.

Real-time PCR quantification will be performed in order to determine differential expression of candidate genes. Primers will be custom-designed using VectorNTI software (107) and will be weighted to the 3' UTR of candidate genes. Colonic mucosal tissue (100mg) from UC, CD, and normal mucosa will be subjected to total RNA

extraction following homogenization and cell lysis in guanidine isothiocyanate-containing buffer using an affinity column technique (Qiagen RNeasy[®] Midi purification kit, CA). Collected cellular RNAs will then be converted to cDNA. All reactions will be performed on an ABI Prism[®] 7700 Sequence Detection System using Stratagene[®] Brilliant[®] Quantitative PCR kit with *β-actin* as an internal standard. Each reaction will be performed in triplicate and reactions without cDNA will be included as negative controls. Cycle threshold (Ct) values will be obtained graphically for each gene of interest and the *β-actin* internal standard. Standard curves will be constructed for all genes before calculation of gene-specific mRNA levels, and gene expression normalized to *β-actin*. For each sample, the mean of three Ct values will be calculated. Comparison of gene expression between control and disease samples is derived from the standard curve method. Relative gene expression will be compared to controls set as normal.

Complete methods used for RNA extraction, cDNA synthesis, and real-time PCR quantification of candidate gene expression are given in Chapter VII.

Polymerase Chain Reaction Primer Design and Amplification Procedure

Candidate gene sequence data are obtained from the National Center for Biotechnology Information (NCBI) human genome public database (<http://www.ncbi.nlm.nih.gov/>). Sequence data are then imported into VectorNTI software (107) for PCR amplicon design. PCR protocols are derived from amplicon-specific analyses performed using VectorNTI software (107), with touchdown thermocycling protocols typically being employed. The primary focus of this part of the

study will be on SNP discovery within candidate gene exons and regulatory regions including the promoter and 3'-untranslated region (3'-UTR). Amplicons for candidate gene exons will include approximately 100bp of intronic DNA on both the 5' and 3' sides of coding regions. These intronic regions will be included since polymorphisms here may be involved in binding of regulatory proteins (e.g. transcriptional enhancers), or may alter mRNA alternative splicing patterns.

Germline DNA extracted from peripheral leucocytes will form the template for PCR amplification of candidate genes. Specific details of PCR methodology will be described in Chapter VII.

Candidate Gene Polymorphism Discovery

Overview: Polymorphism discovery is an integral part of this part of the study. Although gene polymorphism data within human genome databases (e.g. the NCBI dbSNP databases) is useful, it does have limitations. Many of the described polymorphisms have been characterized in a small number of individuals (often from fewer than 20 chromosomes). It is likely that the frequency of disease-associated mutations within susceptibility genes would be low in healthy populations such as those used in the Human Genome Project. It is extremely unlikely that a disease-associated polymorphism would be detected in such a low number of unaffected individuals. Screening for polymorphisms plays an important role given the low level of polymorphism characterization, especially within the proposed candidate genes.

Candidate Gene Amplicon Design: Candidate gene sequence data are obtained from public databases such as the National Center for Biotechnology Information web page (<http://www.ncbi.nlm.nih.gov/>). Sequence data are then imported into VectorNTI software (107) for PCR amplicon design. A number of amplicon design parameters are set, and are based around optimal denaturing high performance liquid chromatography (dHPLC [Transgenomic WAVE[®], Omaha, NE]) amplicon analysis. Desired amplicon characteristics are as follows:

- 1) Ideal amplicon size is between 200-700bp. Sizes of up to 1500bp can be adequately analyzed. Analysis of amplicons <200bp is not feasible.
- 2) Ideal amplicon GC content is between 48-68% GC. Fragments with a higher or lower GC content can be used, but will require dHPLC heteroduplexing temperature optimization.

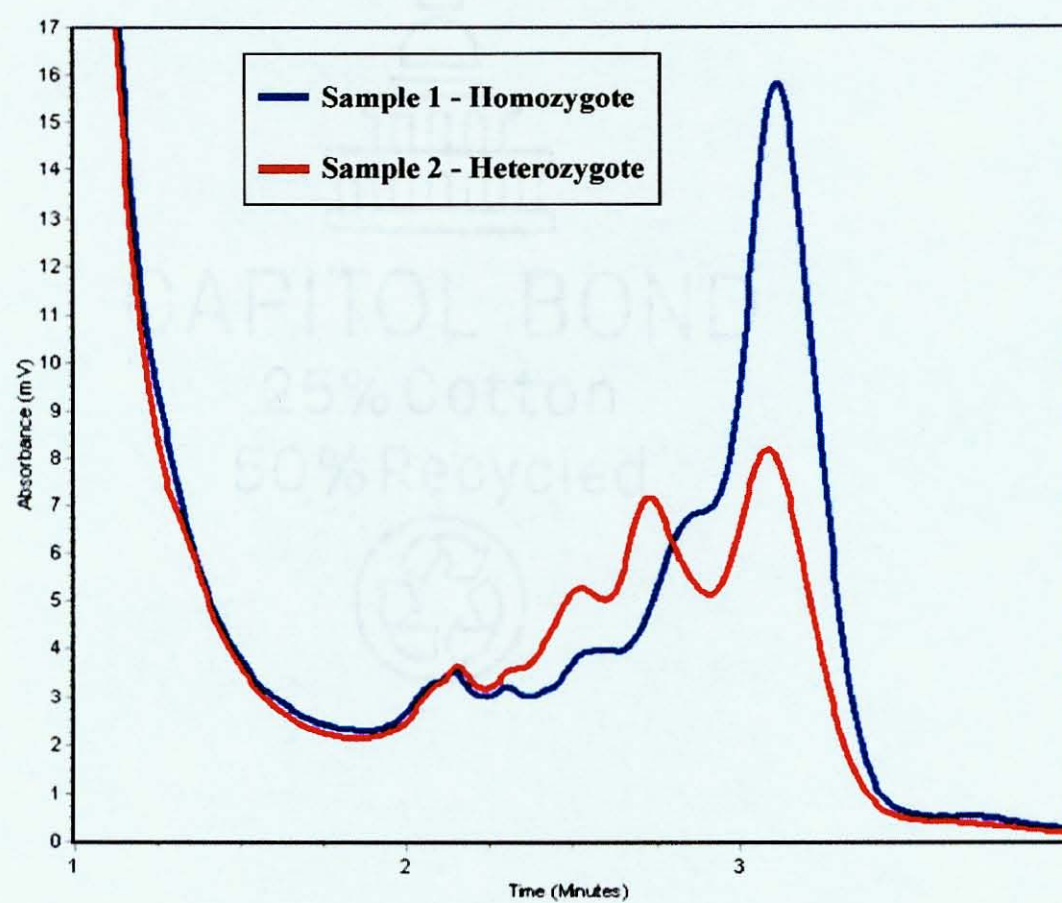
Mutation Analysis: PCR amplicons will be screened for mutations using dHPLC. This method compares two or more chromosomes as a mixture of denatured and re-annealed PCR amplicons. A sample dHPLC chromatogram is shown in Figure 5.

The presence of a polymorphism within one of the two genomic copies of an amplicon (i.e. when an individual is heterozygotic for a polymorphism) is indicated by differential retention of hetero- and homoduplex DNA on reversed-phase chromatography supports under conditions of partial denaturation (108). DNA fragments are eluted from the stationary phase of the HPLC column (DNASep[®] cartridge, Transgenomic, Omaha, NE) by creating a buffer gradient. Buffers used in this process are 0.1M triethylammonium acetate (TEAA) solution ('Buffer A') and 0.1M TEAA containing 25% acetonitrile ('Buffer B'). Acetonitrile interferes with the interaction

Figure 5. Example of a polymorphic and non-polymorphic dHPLC chromatogram used for screening candidate gene amplicons for polymorphisms. Chromatograms from two separate samples are shown here:

Sample 1: One major peak evident. This pattern is seen when there is no amplicon heteroduplexing. The candidate gene amplicon from this individual is not polymorphic.

Sample 2: Two major peaks evident. This pattern is seen when there is amplicon heteroduplexing. Heteroduplexed amplicons are eluted from the dHPLC column earlier than the homoduplexes. These data imply that there is a polymorphism within that amplicon, although it does not confer the position or nature (i.e. insertion, deletion, missense) of the polymorphism. Sequence analysis of the polymorphic amplicon must now be performed.



between the negatively charged DNA molecule, the positively charged TEAA ion-pairing reagent, and the hydrophobic stationary phase of the DNASep® Cartridge. Thus as acetonitrile concentration (i.e. buffer B) increases with time, differentially sized DNA molecules are eluted. Hetero- and homoduplex DNA are resolved in this manner, with heteroduplexes being eluted from the column at an earlier time period. Single nucleotide substitutions, deletions, and insertions can be detected by on-line UV monitoring within 2-3 minutes in unpurified amplicons as large as 1.5kbp. The sensitivity and specificity of this method consistently approaches 100% (108).

The process of heteroduplex resolution is very temperature sensitive. Optimal temperature for mutation analysis is dependant on the helicity of the amplicon being studied. Optimal helicity for dHPLC mutation analysis is between 50% and 98% i.e. 50-98% of the amplicon is double stranded at the experimental temperature. The helicity of an amplicon at a given temperature is determined by its G-C content. Different parts of an amplicon will likely have different G-C contents, and owing to the higher stability of G-C as compared to A-T bonds, higher G-C content amplicon domains will require more energy to disrupt nucleotide pair hydrogen bonding than lower G-C domains. Thus the helicity of G-C rich amplicon domains will be higher than that of A-T rich domains at increasing temperatures. Any given amplicon can therefore be divided into domains based on its G-C content. The location of these domains within an amplicon, and the temperature at which dHPLC experiments must be performed in order for amplicons to have optimal helicity, are predicted by Transgenomic Wavemaker software (109). It is therefore highly probable that an amplicon will have to be analyzed at a variety of temperatures to effectively screen all domains for polymorphisms.

Screening Protocol: A protocol whereby a limited, although representative proportion of the total population are screened for polymorphisms is required for a number of reasons. Firstly, although the total study population consists of over 1000 individuals, many study participants have limited amounts of stock genetic material available. Secondly, the length of time required to screen each of the numerous candidate gene amplicons would be considerable if performed for every member of study cohort. The ‘screening’ cohort will therefore consist of 96 individuals comprising a representative proportion of the total study population (32 population controls, 32 ulcerative colitis, and 32 Crohn’s disease individuals). Every attempt will be made to include individuals in the screening process that have disease-associated STR polymorphisms. Whether this is possible will depend on the amount of stock DNA available for such patients. Ideally, a large quantity of stock genetic material will be available for individuals included in the screening stage of the study.

There are two possible outcomes from dHPLC screening:

- 1) A polymorphism is identified within a candidate gene amplicon. In this instance, the SNP is likely to be of a high enough frequency in the entire population to warrant further study. DNA sequencing of apparently polymorphic amplicons will therefore need to be performed.
- 2) There may be no evidence of polymorphism within an amplicon in the screening population. This will probably be due to the fact that the studied amplicon truly does not contain a polymorphism, or because a polymorphism is too rare to be detected (i.e. likely total population frequency <1%). In the instance of an

infrequently occurring polymorphism, further analysis would not be warranted since it is unlikely to make a significant contribution to disease susceptibility.

DNA Sequencing: Following dHPLC screening of candidate genes, apparently polymorphic amplicons will be sequenced. To ensure complete characterization of wildtype and mutant alleles, DNA sequencing will be performed on several individuals with both polymorphic and non-polymorphic candidate gene amplicons. DNA fragments are sequenced using fluorescent cycle sequencing and analyzed on a DNA sequencer (model CEQ2000XL; Beckman-Coulter, Fullerton, CA). These data will in turn, be used to design fluorogenic probes for SNP-specific PCR, as described below.

Given current data, the number of SNPs in the human genome likely exceeds 3 million (or 1 SNP/kbp) (110). It is therefore highly likely that many polymorphisms will be identified in both candidate genes through the use of this screening process.

SNP Genotyping

Following polymorphism discovery, TaqMan-based genotyping methods will be employed to characterize polymorphisms in the entire study population.

PCR Primer and Oligonucleotide Probe Design: Candidate gene SNP-specific PCR primers and fluorogenic probes will be designed using Primer Express (version 1.5, Applied Biosystems, Foster City, CA). The fluorogenic minor groove binder (MGB) TaqMan probes are labeled with a reporter dye (either FAM [5-(&6)-carboxyfluorescein] or VIC® [a proprietary fluorescent dye produced by Applied Biosystems]) specific for the

wildtype and variant alleles of each candidate gene SNP. For consistency, FAM will always label the wildtype probe and VIC the mutant probe.

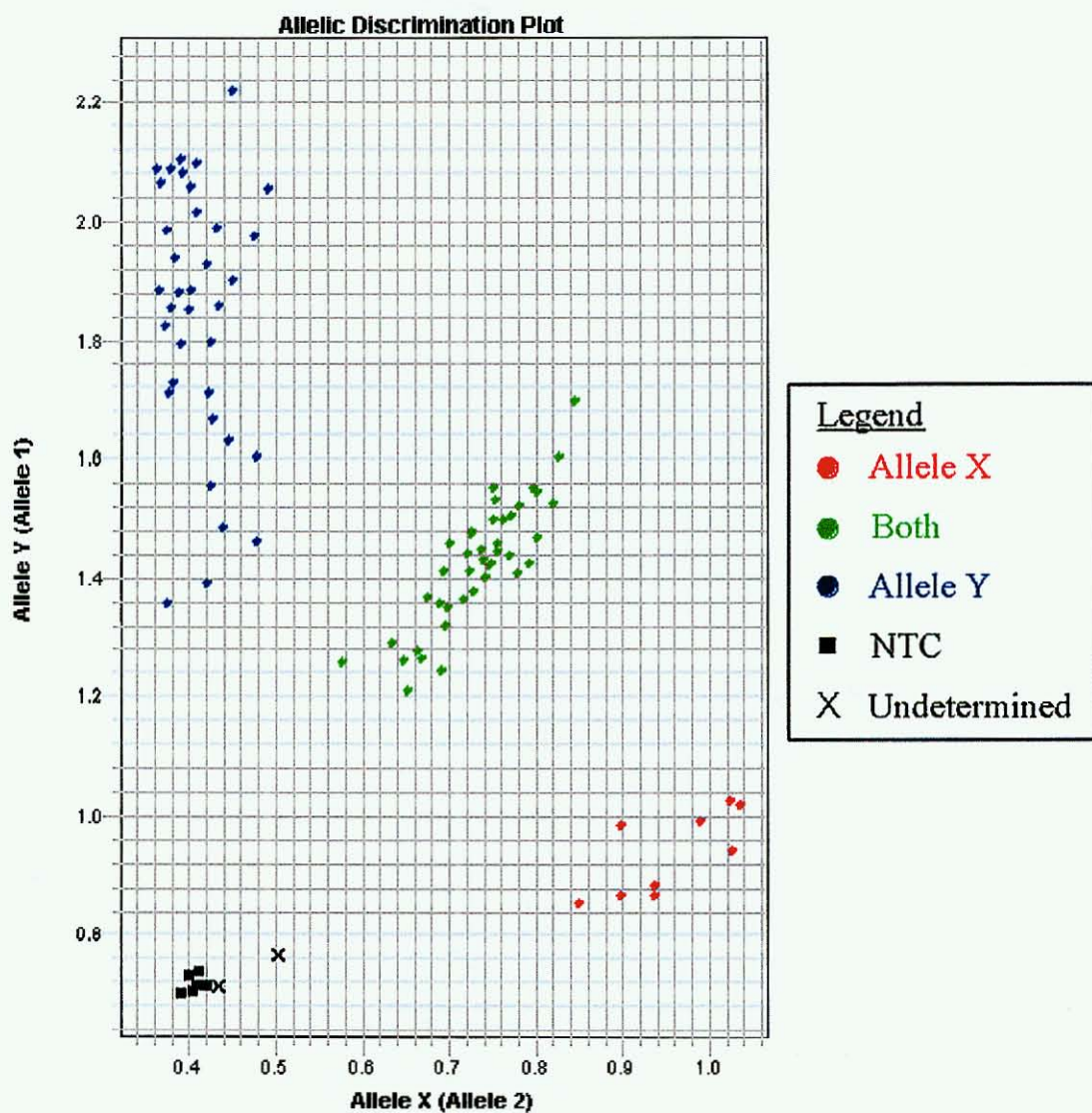
Fluorogenic probe and PCR primer design have been described elsewhere (111). Briefly, fluorogenic probes are designed to have a melting temperature between 65°C and 67°C, and with the SNP located in the center of the oligonucleotide. Melting temperatures of each SNP-specific probe-pair will not differ by more than 1°C. PCR primers will have a melting temperature between 58°C and 60°C and will contain less than three G's and C's within the last five bases at their 3'-end.

PCR Amplification and SNP-Genotyping: PCRs will be carried out using a two temperature PCR protocol (Perkin-Elmer Thermocycler, Norwalk, CT). Reactions lacking template DNA will be included as controls to ensure that there is no amplification of contaminating DNA. The fluorescence of each sample will be read post-PCR on an ABI Prism[®] 7900HT sequence detection system (Applied Biosystems, Foster City, CA). An example of the type of output given by this apparatus is given in Figure 6. Fluorescence is detected when the probe with the exact sequence match binds to the complementary region of the single-stranded DNA template and is digested by the 5'-3'-exonuclease activity of *AmpliTa**q* Gold DNA polymerase (Applied Biosystems, Foster City, CA). Probe digestion releases the fluorescent reporter dye from the quencher located on the 3' end of the probe. Homozygote wildtype samples fluoresce with FAM, whereas homozygote mutants fluoresce with VIC, and heterozygotes with both FAM and VIC.

PCR primers and SNP-specific probe sequences for each SNP included in the study are shown in Chapter VIII.

Figure 6. Post-PCR determination of fluorescence from SNP-specific PCR reactions – typical read-out. Following PCR amplification of a polymorphism, the ABI 7900HT apparatus determines the relative fluorescence of the 6-FAM and VIC marker dyes in each sample. Each data point represents the fluorescence from one reaction. Five groups are apparent:

- 1) **Allele X** – These samples only fluoresce with the allele X dye. By convention, this is always VIC, the dye bound to mutant probes. These individuals are therefore homozygote mutants for the SNP of interest.
- 2) **Allele Y** – These samples only fluoresce with the allele Y dye, 6-FAM. These individuals are therefore homozygote wildtypes for the studied SNP.
- 3) **Both** – Both marker dyes fluoresce with approximately equal intensity. These individuals are heterozygotes for the mutant SNP allele.
- 4) **NTC** – ‘No Template Control’. These are negative controls that contain all reagents apart from template DNA. They control against the presence of contaminating (e.g. microbial) DNA. As expected, fluorescence is not detected for either dye in the samples shown here.
- 5) **Undetermined** – These are either samples where PCR amplification has failed or the position of the data point on the graph is ambiguous. A number of failed reactions are shown here.



Statistical Analysis

Case-Control Association Studies: A case-control approach will be employed for comparison of candidate gene SNP frequencies between IBD populations and the control population. Comparisons will be made using the using chi-squared and/or Fishers Exact tests with use of the Bonferroni method to correct for compounding of type 1 when appropriate. Analyses will be performed in racially matched disease and control populations. Case-control analyses will initially be performed using CD, UC, and IC groups.

Other Statistical Analysis: Haplotype and family-based methods will be used in a similar manner for candidate gene SNPs as were used for STR analysis (see p. 45-47). Linkage disequilibrium is far more likely to exist between SNPs within the same candidate gene owing to their likely closer genetic proximity. It is therefore likely that the use of haplotype estimations will be far more extensive in candidate gene studies than was the case for STR data.

Analysis of Genotype-Phenotype Correlations

Subset analysis will form an essential part of this proposal. The importance of such analyses has been underlined by experience with the IBD susceptibility gene *NOD2*. *NOD2* disease-causing mutations have been shown to be associated with fibrostenosing ileocolic CD in Caucasian populations (31;33;35;36). Recent data from these as well as other association studies conclusively demonstrate that IBD consists of more than the two traditionally described phenotypes (i.e. CD and UC) (32). Analysis of genotype-phenotype relationships in IBD will likely improve statistical power by increasing group

homogeneity. In a similar manner, ethnic differences are another important factor that must be taken into consideration. For example, *NOD2* mutations play a prominent role in CD susceptibility in Caucasians, but not in Japanese populations (38;112).

The CD and UC groups will therefore be subdivided on the basis of ethnicity, gender, age of diagnosis and presence of extra-intestinal disease manifestations. The IC population will not be subdivided in such a manner owing to the low number of subjects currently in the study (n=82). Analyses performed at the sub-phenotype level in such a population would lack statistical power and likely be of little value. The CD population will also be subdivided on the basis of the Vienna classification (98) in order to permit disease location and disease behavior to be considered as distinct CD subgroups.

Following subgroup division, the methods described above (i.e. case-control, haplotype and family-based analyses) will be employed to analyze genotype-phenotype relationships. Each phenotypic variable will be considered separately and analyzed in isolation to other disease characteristics.

CHAPTER V

CHARACTERIZATION OF GENOTYPE-PHENOTYPE RELATIONSHIPS FOR INFLAMMATORY BOWEL DISEASE SUSCEPTIBILITY LOCI USING MULTIPLE SHORT TANDEM REPEAT GENETIC MARKERS

Overview

BACKGROUND: The genetic basis of IBD is well founded and susceptibility loci have been identified on a number of different chromosomes. Previous genome-wide scans have produced highly variable results, with reproduction of linkage demonstrated at some loci proving difficult or impossible. The most likely sources of these variable outcomes are intrinsic population differences and inconsistent IBD phenotyping as a result of differing diagnostic criteria. **HYPOTHESIS:** A variable number of previously described IBD susceptibility loci will be associated with disease in this population.

METHODS: The aim of this study was to improve disease characterization by defining genotype-phenotype relationships for various IBD susceptibility loci (IBD1, IBD2, IBD5-7, chromosome 4) by characterizing previously described peak logarithm of odds (LOD) score short tandem repeat markers. The study population consisted of 484 severely affected Caucasian patients with IBD, 144 unrelated population controls, and 348 non-affected first-degree relatives of IBD patients. Markers were amplified from peripheral leukocyte DNA using polymerase chain reaction and analyzed using high-performance liquid chromatography. Case-control association studies were performed

using the Mehta-Patel extension of Fisher's exact test. Family-based association studies were used to confirm associations from case-control tests. **RESULTS:** IBD1, IBD2, and IBD5 markers were associated with disease in the study population. Following correction for multiple analyses, association was evident in terminal ileal CD for IBD1 ($P=0.012$), and in indeterminate colitis for IBD5 ($P=0.008$). There was also a trend towards significant association in colonic CD for IBD1 ($P=0.022$), in both indeterminate colitis ($P=0.027$) and terminal ileal CD ($P=0.014$) populations for IBD2, and finally in colonic CD for IBD5 ($P=0.017$). Family-based studies confirmed IBD2 association in 13 indeterminate colitis families ($P=0.005$) as well as in 35 ulcerative colitis families ($P=0.048$). **CONCLUSIONS:** These results show that IBD1, IBD2, and IBD5 play a role in disease susceptibility in this IBD population.

Introduction

As discussed in Chapter II, both UC and CD are complex, polygenic disorders that share some, but not all, susceptibility genes (10). Genome-wide linkage studies using non-informative STR markers have defined chromosomal regions where the presence of a certain marker/loci increases an individual's risk of developing IBD. Thus far, nine such IBD susceptibility loci with a relatively high likelihood of linkage to IBD have been identified. These loci have subsequently been named IBD1-9 (5;18-25). In addition to IBD1-9, a number of studies have produced evidence for the existence of other loci, including a chromosome 4 susceptibility locus (23).

Replication of linkage exhibited by the majority of IBD susceptibility loci has been relatively inconsistent. Only IBD1-4 have shown any level of consistent

reproducibility (5;113-116). Confirmation of linkage at other IBD susceptibility loci has often only proved possible within small, homogenous populations. An example of this type of phenomenon is the finding of association with the IBD7 locus in an American Chaldean IBD population (93).

The aim of this study is to characterize IBD susceptibility loci in a phenotypically well-defined rural population through the use of previously described STR markers. The studied STR markers possess peak LOD scores with respect to previously described susceptibility loci. The overall goal of this study was to define IBD genotype-phenotype correlations that could eventually lead to improved disease classification, as well as to define disease-associated genomic regions in the study population as a means to commence more detailed mapping and candidate gene studies.

Materials and Methods

Study Design

Study design has been described in detail in Chapter IV. Briefly, the primary approach was to perform a case-control association study using STR markers with the highest LOD score with respect to previously described IBD susceptibility loci. STR allele and genotype frequencies were compared between well-defined disease populations and an unrelated control population with no clinical evidence or family history of IBD. The CD population was subdivided into populations with distinct phenotypes based on the Vienna classification (98).

In addition to characterizing peak LOD score STR markers, a non-IBD STR from a genomic region that has never displayed linkage to IBD was defined. If associations were evident with this marker, it would imply that associations observed with IBD markers might be spurious. Conversely, if no associations were observed with this “non-IBD” STR, more strength would be afforded to the argument that any association observed with IBD susceptibility loci STRs were valid (117).

Family-based tests of association were performed to confirm the findings of case-control studies, in addition to accounting for the possible confounding effects of population substructure. Family-based tests were performed using non-affected first-degree relatives of IBD probands.

Setting and Participants

The University of Louisville Institutional Review Board approved this study. Written informed consent was obtained from all subjects. Patients were derived from a University-based colon and rectal surgery practice. These patients were referred for surgical consultation by their treating gastroenterologists due to the severity of their symptoms or lack of response to medical therapy, and are therefore classified as having severe disease. Patients originated from a relatively small rural geographical area consisting of Kentucky and southern Indiana. This is atypical since IBD is more prevalent in urban populations (13).

The patient population consisted of a total of 628 Caucasians, including 254 unrelated individuals with CD (63% women), 165 with UC (53% women), 65 with IC

(70% women), and 144 population controls (74% women). About one third of the IBD patients studied did not require surgery (25% of CD patients, 33% of UC patients). The average duration of clinical follow up for the CD population was 4.6 years (range, 0-16 years), 4.3 years for the UC population (range, 0-14 years), and 5.6 years for the IC population (range, 0-12 years).

For the purposes of more accurate phenotyping, the CD group was subdivided based on the Vienna classification (98). This classification categorizes CD patients based on their age of onset (A), location of disease (L), and disease behavior (B). With regard to age of onset, 175 of 254 (69%) CD patients were diagnosed at < 40 years of age (A1 group), 47 of 254 (18%) at \geq 40 years old (A2 group), and information regarding age of onset was not available for the remaining 32 patients (13%). Regarding disease location, 70 of 254 (28%) had terminal ileal disease (L1 group), 152 of 254 (60%) had purely colonic and/or terminal ileal (ileal and colonic) disease (L2/L3 groups), and 32 of 254 (12%) had disease located proximal to the terminal ileum (L4 group). The L2 and L3 groups were combined in this study since colonic disease constitutes the primary focus of research in this cohort. Due to the relatively low numbers of individuals with CD proximal to the terminal ileum (L4 group), this disease subgroup was excluded from case-control analyses. Regarding disease behavior, 96 of 254 patients (38%) had uncomplicated inflammatory disease (B1 group), 28 of 254 (11%) had stricturing disease (B2 group), and 130 of 254 (51%) had penetrating disease (B3 group). This pattern of disease behavior is consistent with the severe affection status of this patient cohort.

Three hundred forty-eight non-affected first-degree relatives of IBD patients were available for this study. There were 63 CD families (29 triads, 30 discordant sibling pairs [DSPs]), 4 triad/DSP combinations), 43 UC families (11 triads, 24 DSPs, 8 triad/DSP combinations), and 14 IC families (5 triads, 5 DSPs, 4 triad/DSP combinations). In addition, there were 17 CD families, 18 UC families, and 9 IC families who could not be classified as either a triad or DSP. CD families included probands with ileal disease, probands with colonic CD, as well as probands with proximal small bowel disease.

Genetic Marker Characterization

STR markers with the highest LOD scores were selected from the original studies describing the IBD susceptibility loci IBD1, IBD2, IBD6, IBD7, and the chromosome 4 susceptibility locus (5;18;19;22;23). The IBD5 marker D5S1984 was chosen from a fine-mapping study (118) of the 5q31 cytokine cluster performed by the group that initially described IBD5. The non-IBD STR marker D13S252, which is located in the pericentromeric region of chromosome 13, was selected from the NCBI human genome database (<http://www.ncbi.nlm.nih.gov>). The genetic markers used in this study and the associated polymerase chain reaction (PCR) primers are shown in Table 1.

PCR primers were custom-synthesized (Proligo, La Jolla, CA) and STRs were amplified from genomic DNA using a standard PCR protocol (Perkin-Elmer Thermocycler, Norwalk, CT). PCR mixtures consisted of 100ng genomic DNA, 70nM of each dNTP, 2.5µl Thermophilic DNA 10x reaction buffer (Promega, Madison, WI), 1.5µM magnesium chloride, 1.0µM of both forward and reverse primers, and 1 unit Taq

Locus	STR Marker	LOD Score	PCR	PCR Primer Sequence
			Annealing Temperature	
IBD1	D16S419	3.17	57°C	ATTTTAAAGGAATGTAAAGNACACA GACGTTAGACCAGGAGTCAG
IBD2	D12S83	5.47	59°C	TTTTTGGAAGTCTATCAATTTGA TAGCAGAGAAAGCCAATTCA
IBD5	D5S1984	3.00	68°C	CCAGCCCGCTTAGTGT TAGGAGGCTTCCCACATCT
IBD6	D19S1166	4.60	54°C	TTCCAGCCTAGGTAGCAGTG GCAACTGAGGAAATGCATCT
IBD7	D1S1597	3.01	55°C	TTTATTGAGATATATTTGACATGCA AAGGAGGAAAGCTTTTTGGA
Chr. 4	D4S1647	2.76	59°C	TATTTCCAACACCCCTGCTA AAGCAAAGAGGATTGAAAGTG
Non-IBD	D13S252	N/A	57°C	CTCAACCTGGCTTCTACT TACTCCTTAATAAACTCCCC

Table 1. STR markers utilized in the IBD susceptibility loci study. Amplimer sequences, LOD score from previous linkage studies and PCR annealing temperatures are shown for each STR marker.

DNA polymerase (Promega, Madison, WI) in a final 25 μ l reaction volume. Thermal cycling parameters comprised an initial denaturation step of 94°C for 10 min followed by 25 cycles of denaturing at 94°C for 30 sec, annealing at the primer-specific temperature shown in Table 1 for 30 sec, extension at 72°C for 30 sec, followed by a final extension phase at 72°C for 7 min. PCR products were analyzed using HPLC (WAVE[®], Transgenomic, Omaha, NE).

Statistical Analysis

Methods for statistical analysis of STR data have been described in detail in Chapter IV. Case-control and family-based approaches have been used here. Individuals with CD proximal to the terminal ileum were excluded from case-control analyses owing to their relatively low numbers in the study population (L4 group, n=32).

To account for multiple testing in case-control studies, individual STR genotype and allele frequencies were only examined when the global test of homogeneity of all groups and disease versus control pair-wise comparisons achieved statistical significance. Global tests of homogeneity of disease versus control pair-wise comparisons were corrected by the Bonferroni method, and were considered to have achieved statistical significance when $P \leq 0.0125$.

With regards to family-based testing, CD pedigrees were not analyzed based on disease location. All CD family data were pooled in order to maximize statistical power by allowing for the analysis of the maximum number of CD families. Haplotype analysis of these data was not applicable.

Results

Allele and Genotype Frequencies

Allele frequencies for STRs that were disease-associated on case-control analysis are shown in Table 2. Smaller alleles were eluted from the DNasep cartridge (Transgenomic, Omaha, NE) earlier than larger alleles, and therefore were detected by the HPLC apparatus at earlier time intervals. Alleles were named on the basis of their size, with “1” representing the largest amplified fragment and consecutive numbers representing alleles of decreasing size.

Case-Control Association Studies

Results of global case-control analyses for all STRs are shown in Table 3. These P-values are uncorrected for multiple analyses. Each locus is now considered in turn.

IBD1: The global test of homogeneity of all groups for the IBD1 STR (D16S419) alleles showed a statistical difference ($P=0.004$). Following this, the global test of homogeneity between disease and control groups demonstrated a statistically significant difference in allele frequencies between the ileal (L1) CD groups when compared with the control population ($P=0.012$). This was significant at the Bonferroni-corrected significance level of $P<0.0125$). Individual pair-wise disease *versus* control analyses showed over-representation of allele 5 in the ileal (L1) CD group (ileal CD frequency 10 of 50 alleles [20%], control frequency 5 of 118 alleles [4%], $P=0.002$). The uncorrected global test of homogeneity between disease and control groups demonstrated a statistically significant difference in allele frequencies between colonic (L2/L3) CD groups when compared with

IBD1 D16S419 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
4	53	33%	28	33%	14	28%	31	28%	49	42%
5	11	7%	8	10%	10	20%	18	16%	5	4%
7	9	6%	2	2%	1	2%	9	8%	10	8%
6	71	44%	36	43%	19	38%	42	38%	44	37%
Pooled	16	10%	10	12%	6	12%	10	9%	10	8%
	160		84		50		110		118	

IBD2 D12S83 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
7	78	43%	33	43%	25	35%	57	38%	34	28%
8	24	13%	7	9%	3	4%	16	11%	21	18%
9	16	9%	14	18%	12	17%	20	13%	11	9%
10	17	9%	8	11%	6	8%	14	9%	21	18%
11	32	18%	11	14%	13	18%	34	23%	20	17%
Pooled	15	8%	3	4%	13	18%	9	6%	13	11%
	182		76		72		150		120	

IBD5 D5S1984 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
2	14	11%	1	2%	2	4%	6	6%	8	10%
4	54	43%	26	45%	21	40%	47	46%	37	46%
5	39	31%	22	38%	16	31%	36	35%	13	16%
6	11	9%	8	14%	11	21%	8	8%	14	18%
Pooled	8	6%	1	2%	2	4%	5	5%	8	10%
	126		58		52		102		80	

Table 2. STR allele frequencies for IBD1, IBD2, and IBD5. Alleles with <5% frequency were pooled.

STR Marker		Disease vs. Control Comparison Global P-value				Global test of homogeneity (combined P value)
		Ulcerative Colitis	Indeterminate Colitis	Terminal Ileal CD (L1)	Colonic Crohn's Disease (L2/3)	
IBD1	Genotype	0.731	0.236	0.640	0.104	0.348
D16S419	Allele	0.423	0.131	0.012	0.022	0.004
IBD2	Genotype	0.283	0.102	0.502	0.032	0.053
D12S83	Allele	0.090	0.027	0.014	0.051	< 0.001
IBD5	Genotype	0.173	0.108	0.623	0.100	0.095
D5S1984	Allele	0.082	0.008	0.159	0.017	< 0.001
IBD6	Genotype	0.416	0.328	0.620	0.482	0.603
D19S1166	Allele	0.568	0.523	0.596	0.547	0.792
IBD7	Genotype	0.830	0.260	1.000	0.108	0.484
D1S1597	Allele	0.973	0.517	1.000	0.348	0.900
Chr.4	Genotype	0.908	0.414	0.335	0.706	0.774
D4S1647	Allele	0.594	0.247	0.021	0.503	0.115
Non-IBD	Genotype	0.117	0.245	0.049	0.833	0.096
D13S252	Allele	0.107	0.094	0.210	0.291	0.063

Table 3. Case-control association test global P-values for IBD loci peak LOD score

STRs. Statistically significant global tests prior to correction for multiple analysis are highlighted.

the control population ($P=0.022$). This, however, was not significant when applying the Bonferroni method for correction of multiple analyses. Some interesting associations were noted with individual pair-wise disease versus control analyses showed over-representation of allele 5 (colonic CD frequency 18 of 110 alleles [16%], control frequency 5 of 118 alleles [4%], $P=0.003$) and under-representation of allele 4 in the colonic (L2/L3) CD population (colonic CD frequency 31 of 110 alleles [28%] compared to the control frequency 49 of 118 alleles [42%], $P=0.038$). These however can only be viewed as trends towards significance and not firm evidence of association.

IBD5: The global test of homogeneity of all groups for the IBD5 STR (D5S1984) alleles showed a statistical difference ($P<0.001$). Following this, the Bonferroni-corrected global test of homogeneity between disease and control groups revealed a significant difference in allele frequencies between the IC groups when compared with the control population ($P=0.008$). Individual pair-wise disease *versus* control analyses showed over-representation of allele 5 in the IC population (IC frequency 22 of 58 alleles [38%], control frequency 13 of 80 alleles [16%], $P=0.005$).

In a similar manner to IBD1, the uncorrected global test of homogeneity between disease and control groups demonstrated a significant difference in allele frequencies between colonic (L2/L3) CD groups when compared with the control population ($P=0.017$). This, however, was not significant when applying the Bonferroni method for correction of multiple analyses, and can consequently only be viewed as a trend towards association. Individual pair-wise disease *versus* control analyses showed over-representation of allele 5 in the colonic (L2/L3) CD group (colonic CD frequency 36 of 102 alleles [35%], control frequency 13 of 80 alleles [16%], $P=0.004$).

IBD2: The global test of homogeneity of all groups for the IBD2 STR (D12S83) alleles was statistically significant ($P < 0.001$). Following this, the global test of homogeneity between disease and control groups revealed a statistically significant difference in allele frequencies between both IC and ileal (L1 group) CD when compared with the control population ($P = 0.027$ and $P = 0.014$, respectively). However, following application of the Bonferroni correction neither P-value was significant. Individual pair-wise disease *versus* control analyses showed over-representation of allele 7 in the IC group (IC frequency 33 of 76 alleles [43%], control frequency 34 of 120 alleles [28%], $P = 0.032$) and under-representation of allele 8 in the ileal CD population (ileal CD frequency 3 of 72 alleles [4%] *versus* control frequency 21 of 120 alleles [18%], $P = 0.006$).

Other markers: Case-control analyses for the IBD6, IBD7, and chromosome 4 susceptibility loci markers did not reveal any evidence of association. As expected, no associations were observed with the “non-IBD” STR in any disease group.

Disease Behavior and Age of Onset

With regard to CD disease behavior characteristics, no correlation was evident between disease behavior and STR allele and/or genotype data. When analyzing age of onset, however, the IBD1 STR (D16S419) global test of genotype homogeneity showed a significant difference in distribution ($P = 0.032$). Following this, the global test of homogeneity between different age of onset groups revealed a significant difference in the age of onset of UC with respect to D16S419 genotype ($P = 0.011$). Individual pair-wise analyses of the 80 UC individuals who had both age of onset and genotype data showed

over-representation of genotype 6,6 in those UC patients with an age of disease onset \geq 40 years (9 of 20 individuals [45%]) compared to those with an age of onset of < 40 years (8 of 60 individuals [13%]; $P=0.009$).

Family-Based Association Studies

Likelihood ratio testing showed evidence of association for D12S83 in 13 IC families ($\chi^2=16.6$ [6 d.f.]; $P=0.005$). This was primarily due to over-transmission of allele 11 ($P=0.022$). The same STR revealed evidence of association in 35 UC families ($\chi^2=12.4$ [8 d.f.]; $P=0.048$). Correction for multiple analyses was not performed. Other STRs did not exhibit evidence of association upon family-based testing.

Discussion

This study presents evidence for disease associations with peak LOD score STRs that have been shown to be linked to the IBD susceptibility loci IBD1, IBD2, and IBD5. Conversely, no association was evident in this population for IBD6, IBD7, and the chromosome 4 susceptibility loci. This is not entirely surprising, since there was a relatively low level of replication of linkage for the latter three susceptibility loci. It may well be that such susceptibility loci have either a limited effect on disease susceptibility, or their effects may be confined to isolated populations. Additionally, the “non-IBD” marker D13S252 did not demonstrate any disease group associations in this population, adding strength to the argument that the association exhibited by the IBD1, IBD2, and IBD5 markers is not spurious.

This study has shown IBD1 to be associated with terminal ileal CD. This result is not entirely surprising given that this finding is undoubtedly due to the presence of *NOD2* within IBD1. Gasche et al. (32) reviewed genotype-phenotype correlations for the IBD1 susceptibility gene *NOD2* and concluded that mutations within this gene were associated with A1/L1 or L3/B2 subgroups of the Vienna classification. These analyses confirm this observation with respect to disease location. Specifically, an association with L1 and L2/L3 CD is demonstrated. This is an important observation given the other STR disease associations that have been found in the current study. There was also some evidence of association with colonic CD (Vienna groups L2/L3). However, following correction for multiple analyses, this association appeared of uncertain significance. It should be noted that the current study has low numbers of participants and thus lacks somewhat in statistical power. The global comparison between colonic CD versus controls for IBD1 alleles approached the significance threshold of $P \leq 0.0125$, with the estimation of the global exact test being $P = 0.022$. With increased numbers, and consequently increased power, it is not improbable that this global test would become significant.

An obvious criticism of this current study is that the number of individuals included is too low to provide convincing evidence of association, especially given the likely modest effects that would be observed with the type of genetic marker studied. Using the IBD1-linked STR D16S419, replication of the association observed in distinct phenotypic groups with *NOD2* has been achieved in this relatively small but clinically homogenous population. It can therefore be argued that the additional case-control associations with IBD5, and to a lesser extent IBD2, are likely true in this cohort.

This study provides convincing evidence to show that IBD5 is associated with indeterminate colitis, with the global test of homogeneity of alleles for IC versus controls being significant ($P=0.008$). This appears to be a novel finding, the likes of which has not been described in other populations. Furthermore, the analyses presented here partially confirm the findings of recent studies with the IBD5 susceptibility locus. Recent studies have demonstrated that the IBD5 locus on chromosome 5q is involved in CD susceptibility (118;119). The region of association has recently been refined to a conserved 250 kb haplotype on chromosome 5q31, which includes the STR studied here D5S1984 (118). Specifically, such studies have shown an association with perianal CD. IBD5 was found to be associated with colonic CD (Vienna classification groups L2/L3). In fact, many of the patients with colonic disease in the study do also have perianal involvement, reflecting the severity of the disease process in this population. This finding should, however be viewed with caution since the global test of allele homogeneity between colonic CD and controls only approached significance ($P=0.017$). This again is likely a reflection of the low statistical power of this study.

Case-control analyses demonstrated evidence of allelic association for the IBD2-linked marker D12S83 in both IC and terminal ileal CD populations. Again, both these associations approached significance following correction for multiple analyses ($P=0.027$ and $P=0.014$ respectively). However, the finding of association in the IC population upon family-based testing gives strength to the argument that case-control studies suffer from low statistical power. That is, the family-based study confirms the apparently non-significant result from case-control analysis. In addition, family based testing showed evidence of association in the UC population, although this was not evident on case-

control analysis. Previous studies showing that IBD2 appears to play a role in the pathogenesis of all forms of IBD are thus confirmed (19;120). A number of novel findings are however presented for the IBD2 susceptibility locus. These include association of various D12S83 alleles with IC and possibly ileal CD. Since the most compelling evidence of IBD2 disease association was displayed in the IC population, it is possible that the IBD2 susceptibility gene plays an important role in the pathogenesis of this form of IBD. Previous studies in the same cohort have shown evidence of genetic heterogeneity between racial and ethnic populations with CD (94). IBD2 therefore appears to be a promising candidate for more detailed characterization by studying multiple STR markers.

Family-based association studies played an integral role in the analytic process. These tests confirmed the presence of association in 13 IC families for the IBD2 STR (D12S83). This result is particularly surprising given the low number of families available for analysis. This implies that the degree of association for this marker may be relatively high in the IC population. CD families were analyzed without accounting for disease location in order to maximize the number of families available. This may explain the lack of confirmation of association observed in case-control studies in the terminal ileal CD population upon family-based testing for D12S83. The proband had terminal ileal CD in only 24 families. Since both the PDT and likelihood ratio tests require larger numbers of informative families to reveal associations, separate analysis of this small group would likely not yield useful data due to low power. Similarly, this probably accounts for the lack of association observed with the IBD1 STR (D16S419) in the CD population upon family-based testing.

A distinct strength of this current study is that a single gastrointestinal pathologist with an interest in IBD reviewed all of the IBD histology. Silverberg et al. (121) suggested that even low rates of diagnostic misclassification could lead to a significant loss of power to detect true linkage, particularly for loci with modest effects, as is likely the case in IBD. Inconsistent diagnoses are likely to account, at least in part, for the contradictory findings of previous linkage studies. It is logical to assume that diagnostic misclassification also would reduce the power to detect association. Minimizing interobserver variation in the diagnosis of IBD should increase the probability that the associations presented herein are true.

The choice of only one STR from each susceptibility locus may be a point of criticism. This is especially true given that previous studies have demonstrated that most susceptibility loci display linkage over broad chromosomal regions. The aim of this study, however, was not to re-define entire susceptibility loci through association studies, but to determine IBD genotype-phenotype relationships, and to define possible starting points for the fine mapping of susceptibility loci. This aim is considered to have been adequately achieved with respect to the IBD2 and IBD5 susceptibility loci. Kaplan and Morris (122) demonstrated that the use of a single marker in an outbred population is a valid approach to define starting points for fine mapping. This provides further evidence to suggest that the genetic region in surrounding the IBD2 STR (D12S83) and the IBD5 STR (D5S1984) will be a useful starting point in this type of study.

The importance of defining IBD genotype-phenotype relationships is becoming increasingly apparent. Descriptions of such relationships will be important in the characterization of the genetic defects that increase susceptibility to the development of

IBD. A number of studies have, for example, correlated the presence of CD-associated *NOD2* mutations in Caucasians with distinct IBD phenotypes (31;33;35;36). Specific *NOD2* mutations have been shown to be associated with a higher incidence of ileal disease, a lower likelihood of colonic involvement, as well as an increased likelihood of developing fistulae and strictures. In a similar manner, IBD1, IBD5, and to some extent IBD2 have been demonstrated to be associated with equally distinct IBD phenotypes.

CHAPTER VI

CHARACTERIZATION OF GENOTYPE-PHENOTYPE RELATIONSHIPS WITHIN THE INFLAMMATORY BOWEL DISEASE SUSCEPTIBILITY LOCUS IBD2

Overview

BACKGROUND: The IBD2 susceptibility locus, initially described by Satsangi et al (1996), spans the majority of chromosome 12, with the peak region of linkage being at 12q13. Subsequent studies showed that IBD2 appears to be more strongly linked to UC than to CD. Although a number of plausible candidate genes have been considered, the identity of the IBD2 susceptibility gene remains unknown. **HYPOTHESIS:** The IBD2 susceptibility locus is associated with disease in the Kentucky and Southern Indiana IBD patient population. **METHODS:** The aim of this study is to perform a detailed characterization of the IBD2 susceptibility locus and to define genotype-phenotype correlations. This will aid characterization of the *IRAK-M* candidate gene that lies within the 12q13 genetic region. The studied population consisted of 484 Caucasian IBD patients, 144 population controls, and 348 non-IBD patients whose first-degree relatives are diagnosed with IBD. Three short tandem repeat (STR) markers were used to characterize the IBD2 locus. Case-control, haplotype, and family-based association tests were used to define genotype-phenotype relationships. **RESULTS:** Using family-based and case-control methods, novel evidence is provided to show that IBD2 is associated

with IC. Family-based testing showed evidence of STR D12S83 allelic association in 13 IC families ($P=0.005$), and evidence of linkage disequilibrium between the D12S1056/D12S83 STRs in the IC family population ($P=0.001$). Case-control analyses partially confirmed these results, with D12S83 alleles being differentially distributed between IC and controls ($P=0.027$). However, this was not significant following correction for multiple analyses. Additionally, following correction for multiple analyses one STR was associated with the development of colonic CD based on the results of case-control studies ($P=0.011$). Family-based analyses of data from one STR also showed association in 38 UC families ($P=0.030$). **CONCLUSION:** The results presented here confirm previous studies showing involvement of IBD2 in susceptibility to all forms of IBD. In contrast to this, IBD2 is demonstrated to be more strongly associated with IC than with UC. This study forms the basis for characterization of the IBD candidate gene *IRAK-M*, which lies in very close genetic proximity to the peak region of linkage within IBD2.

Introduction

The IBD2 susceptibility locus, described by Satsangi et al. (19) in a Northern European cohort, is the focus of the current study. IBD2 spans the majority of chromosome 12 (12p13.2-q24.1). A peak LOD score of 5.47 was obtained with the STR marker D12S83 (located on chromosome 12q13). Parkes et al. (123) further characterized this susceptibility locus and demonstrated linkage heterogeneity between UC and CD. IBD2 was found to have a major role in UC susceptibility, but only a

relatively minor effect in CD. To date, the identity of the IBD2 susceptibility gene remains unknown (124).

The aim of this study was to define the IBD2 susceptibility locus in CD, UC, and IC populations by characterizing the STR marker D12S83, which had the highest LOD score in previous linkage studies, as well as two flanking markers. Disease populations were sub-classified into distinct phenotypic groups in an attempt to ensure clinically homogenous populations. This study was performed in order to define IBD genotype-phenotype correlations, which could eventually lead to improved disease classification and aid future candidate gene studies.

Materials and Methods

Study Design

Study design has been described in detail in Chapter IV. Briefly, a case-control association study was designed in which the IBD2 susceptibility locus was characterized using three STR markers. The principal marker chosen was D12S83, which is the STR marker with the highest LOD score within IBD2 (19). Two further markers were chosen on the basis of their physical and genetic proximity to D12S83 (D12S1056 and D2S1022). STR allele and genotype frequencies were compared between well-defined disease populations and an unrelated control population with no clinical evidence or family history of IBD. Both populations were derived from the same geographic area. The CD population was sub-divided into distinct phenotypic groups based on the Vienna

classification (98), which classifies patients according to their age of onset, disease location, and disease behavior.

Two additional analytical approaches were then applied to the STR data: (1) haplotype estimation and (2) family-based tests of association. Family-based tests enabled confirmation of associations identified in case-control studies, accounting for the effects of population substructure.

Setting and Participants

This study was performed prospectively and had the approval of the University of Louisville Institutional Review Board. Written informed consent was obtained from all subjects. The patient population consisted of a total of 484 Caucasians, including 254 unrelated individuals with CD (63% women), 165 with UC (53% women), and 65 with IC (70% women). For comparative purposes 144 population controls (74% women) and 348 non-IBD patients whose first-degree relatives are diagnosed with IBD were included in the study. The mean patient age was 46 years for UC (range, 11-80 years), 38 years for IC (range, 17-61 years), 40 years for CD (range, 15-81 years), and 55 years for the control population (range, 9-87 years). The mean age of disease onset in UC was 33 years (range, 9-75 years), 30 years in IC (range, 12-59 years), and 30 years in CD (range, 6-81 years). Although the majority of patients had prior surgery, 25% of CD patients and 33% of UC patients were medically managed. The average duration of clinical follow-up for the CD population was 4.6 years (range 0-16 years), 4.3 years for the UC population (range, 0-14 years), and 5.6 years for the IC population (range, 0-12 years).

For the purposes of categorizing disease phenotype more accurately, the CD group was subdivided based on the Vienna classification (98). This classification categorizes CD patients based on age of onset (A), location of disease (L), and disease behavior (B). With regard to age of onset, 175 of 254 (69%) CD patients were diagnosed at < 40 years of age (A1 group) and 47 of 254 (18%) had disease diagnosed at \geq 40 years of age (A2 group). Information regarding age of onset was not available for 32 of 254 patients (13%). Regarding disease location, 70 of 254 (28%) had terminal ileal disease (L1 group), 152 of 254 (60%) had purely colonic and/or terminal ileal (ileal and colonic) disease (L2/L3 groups), and 32 of 254 (12%) had disease located proximal to terminal ileum (L4 group). The L2 and L3 groups were combined for the purposes of this study, since colonic disease constitutes the primary focus of research in this population. Regarding disease behavior, 96 of 254 (38%) had uncomplicated inflammatory disease (B1 group), 28 of 254 (11%) had stricturing disease (B2 group), and 130 of 254 (51%) had penetrating disease (B3 group).

A total of 348 non-affected first-degree relatives of IBD patients were available for the study. There were 63 CD families (29 triads, 30 discordant sibling pairs [DSPs]), 4 triad/DSP combinations), 43 UC families (11 triads, 24 DSPs, 8 triad/DSP combinations), and 14 IC families (5 triads, 5 DSPs, 4 triad/DSP combinations). In addition, there were 17 CD families, 18 UC families, and 9 IC families who could not be classified as either a triad or DSP. CD families included probands with ileal disease, probands with colonic CD, as well as probands with proximal small bowel disease.

STR Characterization

Two markers in close proximity to D12S83 ([AC]_n) were selected from the Fondation Jean Daussèt database (<http://cephb.fr/cgi-bin/wdb/ceph/systeme/form>). The tetranucleotide STRs D12S1056 (GATA)_n and D12S1022 (GAAT)_n were chosen for study. These markers are 0.35Mbp proximal and 2.70Mbp distal of D12S83, respectively (<http://www.ncbi.nlm.nih.gov>).

Polymerase chain reaction (PCR) primers were custom-synthesized (Proligo, La Jolla, CA) and had the following sequences:

D12S83

5'-TTTTTGGAAGTCTATCAATTTGA-3' and 5'-TAGCAGAGAAAGCCAATTCA-3'

D12S1022

5'-AAGATCAGGTCAGTTGCGAA-3' and 5'-CACTTCAAGGAGCCAGTGGA-3'

D12S1056

5'-CAAATTCAACCATCATCACT-3' and 5'-GCAGCATTGAACATTTAACA-3'

STRs were amplified using a standard PCR protocol (Perkin-Elmer Thermocycler, Norwalk, CT). PCR mixtures for all markers consisted of 100ng of genomic DNA, 70nM of each dNTP, 2.5µl of thermophilic DNA 10x reaction buffer (Promega, Madison, WI), 1.5µM of magnesium chloride, 1.0µM of both forward and reverse primers, and 1 unit of Taq DNA polymerase (Promega, Madison, WI) in a final of 25µl reaction volume. Thermal cycling parameters for STRs comprised an initial denaturation step of 94°C for 10 min, followed by 25 cycles of denaturing at 94°C for 30

sec, annealing at the primer-specific temperature (59°C for D12S83, 58°C for D12S1022, and 55°C for D12S1056) for 30 sec, extension at 72°C for 30 seconds, followed by a final extension phase at 72°C for 7 min.

STR amplicons were analyzed using HPLC (WAVE[®], Transgenomic, Omaha, NE). Two independent observers blinded to disease diagnosis and patient identity determined genotypes.

Statistical Analysis

Statistical analysis methods for STR data have been described in detail in Chapter IV. Case-control and family-based approaches have been used here. Individuals with CD proximal to the terminal ileum were excluded case-control analyses owing to their relatively low numbers in the study population (L4 group, n=32).

To account for multiple testing in case-control studies, individual STR genotype and allele frequencies were only examined when the global test of homogeneity of all groups and the disease versus control pair-wise comparisons achieved statistical significance. The global tests of homogeneity for disease versus control pair-wise comparisons were corrected by the Bonferroni method, and were considered to be statistically significant when $P \leq 0.0125$.

Haplotype analysis was performed using two methods. Firstly, haplotype frequencies were estimated using an expectation maximization method as described in Chapter IV. In addition to this phase unknown approach, haplotype frequencies were estimated using transmission disequilibrium test (TDT) analysis (TRANSMIT [105]).

This approach has the important advantage of correcting for population substructure. Haplotype analyses were performed for D12S83 and D12S1056 STR genotype data (inter-SNP distance = 350kbp), since previous studies have shown that linkage disequilibrium can span several hundred kbp (125). Since D12S1022 was 2.70 Mbp distal to D12S83, haplotype analyses were not performed for these markers.

With regards to family-based testing, CD pedigrees were not analyzed based on disease location and all CD family data were pooled. This maximized statistical power by allowing for the analysis of the maximum number of CD families.

Results

Allele and Genotype Frequencies

Alleles were named numerically according to molecular weight, with “1” representing a larger allele than “2,” etc. Larger PCR fragments were eluted from the DNasep cartridge (WAVE[®], Transgenomic, Omaha NE) less rapidly than smaller alleles. Allele and genotype frequencies for D12S83, D12S1056, and D12S1022 are shown in Tables 4 (alleles) and 5 (genotypes).

Case-Control Association Tests

Results of global case-control analyses for all STRs are shown in Table 6. The P-values shown are uncorrected for multiple analyses.

Table 4. Allele frequencies for the short tandem repeats (STR) D12S83, D12S1056, and D12S1022. * Pooled alleles with frequency <5%

STR D12S83 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
7	78	43%	33	43%	25	35%	57	38%	34	28%
8	24	13%	7	9%	3	4%	16	11%	21	18%
9	16	9%	14	18%	12	17%	20	13%	11	9%
10	17	9%	8	11%	6	8%	14	9%	21	18%
11	32	18%	11	14%	13	18%	34	23%	20	17%
Pooled*	15	8%	3	4%	13	18%	9	6%	13	11%
	182		76		72		150		120	

STR D12S1056 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
3	54	22%	23	25%	28	32%	38	19%	53	23%
4	121	50%	39	42%	32	36%	83	42%	95	41%
5	54	22%	25	27%	23	26%	63	32%	67	29%
Pooled*	15	6%	5	5%	5	6%	14	7%	15	7%
	244		92		88		198		230	

STR D12S1022 Allele	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
3	29	12%	8	8%	7	7%	30	14%	34	14%
4	203	81%	85	82%	82	82%	172	79%	176	75%
Pooled*	18	7%	11	11%	11	11%	16	7%	26	11%
	250		104		100		218		236	

Table 5. Genotype frequencies for the short tandem repeats (STRs) D12S83, D12S1056, and D12S1022. * Pooled genotypes with frequency <5%

STR D12S83 Genotype	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
7,7	17	19%	9	24%	6	17%	15	20%	4	7%
7,8	7	8%	2	5%	0	0%	2	3%	4	7%
7,9	8	9%	6	16%	4	11%	8	11%	4	7%
7,10	9	10%	1	3%	4	11%	3	4%	7	12%
7,11	14	15%	3	8%	3	8%	11	15%	7	12%
8,11	5	5%	2	5%	1	3%	2	3%	5	8%
9,11	4	4%	2	5%	1	3%	7	9%	1	2%
Pooled*	27	30%	13	34%	17	47%	27	36%	28	47%
	91		38		36		75		60	

STR D12S1056 Genotype	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
3,3	6	5%	2	4%	6	14%	3	3%	6	5%
3,4	20	16%	12	26%	11	25%	19	19%	26	23%
3,5	17	14%	6	13%	5	11%	10	10%	15	13%
4,4	36	30%	5	11%	6	14%	19	19%	17	15%
4,5	23	19%	14	30%	5	11%	21	21%	27	23%
5,5	6	5%	2	4%	6	14%	13	13%	9	8%
Pooled*	14	11%	5	11%	5	11%	14	14%	15	13%
	122		46		44		99		115	

STR D12S1022 Genotype	Ulcerative Colitis		Indeterminate Colitis		Terminal Ileal Crohn's Disease (L1)		Colonic Crohn's Disease (L2/3)		Controls	
	N	%	N	%	N	%	N	%	N	%
2,4	8	6%	10	19%	9	18%	11	10%	10	8%
3,4	24	19%	6	12%	6	12%	28	26%	17	14%
4,4	83	66%	34	65%	33	66%	65	60%	72	61%
Pooled*	10	8%	2	4%	2	4%	5	5%	19	16%
	125		52		50		109		118	

STR Marker		Disease vs. Control Comparison Global <i>P</i> value				Global test of Homogeneity (Combined <i>P</i> Value)
		Ulcerative Colitis	Indeterminate Colitis	Terminal Ileal CD (L1)	Colonic CD (L2/3)	
D12S83	Genotype	0.283	0.102	0.502	0.032	0.053
	Allele	0.090	0.027	0.014	0.051	< 0.001
D12S1022	Genotype	0.188	0.024	0.059	0.011	0.001
	Allele	0.179	0.211	0.169	0.371	0.147
D12S1056	Genotype	0.201	0.862	0.188	0.021	0.068
	Allele	0.408	0.977	0.525	0.506	0.810

Table 6. Case-control association global P-values for disease group versus control comparisons for IBD2 STRs.

When case-control analyses were performed for IBD2 STRs, the global test of homogeneity between all groups showed a statistically difference only for D12S1022 genotypes ($P<0.001$) and D12S83 alleles ($P<0.001$).

Following this, global testing of homogeneity between disease and control groups indicated difference in D12S1022 genotype frequencies ($P=0.011$) between colonic CD (L2/L3 groups) and IC when compared to the control population ($P=0.024$). The global test of genotype homogeneity was significant at the Bonferroni-corrected significance level of $P\leq 0.0125$ for the colonic CD versus control comparison. The same cannot be said for the IC population, and as a consequence individual allelic associations in this group should be approached with caution. Individual pair-wise disease versus control analyses showed over-representation of genotype 3,4 in the colonic CD group (colonic CD frequency 28 of 109 genotypes [26%] versus control frequency 17 of 118 genotypes [14%]; $P=0.045$). Despite a significant global P value for the IC versus control genotype comparison, no single causative genotype could be identified.

The global test of homogeneity between disease and control groups indicated a statistically significant difference in D12S83 allele frequencies between both IC and ileal CD (L1 group) when compared to the control population ($P=0.027$ and $P=0.014$ respectively). Following correction for multiple testing using the Bonferroni method, with a correct significance level of $P\leq 0.0125$, neither was significant. Therefore individual allelic associations should be approached with some caution. Individual pair-wise disease versus control analyses showed over-representation of allele 7 in the IC group (IC frequency 33 of 76 alleles [43%], control frequency 34 of 120 alleles [28%]; $P=0.032$) and under-representation of allele 8 in the ileal CD population (ileal CD

frequency 3 of 72 alleles [4%] versus control frequency 21 of 120 alleles [18%]; $P=0.006$).

Family-Based Tests

Family-based association studies were employed to confirm associations identified in case-control analyses. Correction for multiple analyses was not performed. Likelihood ratio testing confirmed evidence of allelic association for D12S83 in 13 IC families (global $\chi^2=14.7$ [4 d.f.]; $P=0.005$). Individual allele distributions among families for this marker are shown in Table 7. These showed significant over-transmission of D12S83 allele 11 in IC families ($P=0.02$).

Significant evidence of association was detected for the STR D12S1056 in 38 UC families based on PDT analysis (global sum PDT $\chi^2=12.8$ [5 d.f.]; $P=0.030$). Individual allele distributions among families for this marker are shown in Table 8. Analysis of individual alleles indicated that this was caused by over-transmission of allele 4 ($P=0.03$) and under-transmission of allele 3 ($P=0.01$). The STR D12S1022 did not identify evidence of association as based on family-based testing.

Haplotype Analysis

TDT analysis of D12S1056/D12S83 haplotypes revealed evidence of linkage disequilibrium between these markers in the IC population. Haplotype frequencies are shown in Table 9. An initial global test of homogeneity showed a statistically significant difference in haplotype frequency in the IC population ($P=0.001$). Analysis of individual haplotype frequencies revealed over-transmission of the haplotype consisting of

Allele*	Observed	Expected	Variance (O-E)	χ^2	<i>P</i>
7	10	11.81	3.90	0.84	0.36
9	6	4.59	1.09	1.82	0.18
10	2	2.52	1.03	0.26	0.61
11	6	3.93	0.81	5.26	0.02

Table 7. Transmission disequilibrium test analysis of D12S83 alleles among indeterminate colitis families. * Only those alleles with a frequency $\geq 5\%$ are shown.

Allele*	Parental Contribution		Discordant Sibs		Z	P
	Transmitted	Not Transmitted	Affected	Unaffected		
2	1	1	1	0	0.82	0.41
3	2	9	7	14	-2.52	0.01
4	17	10	22	25	2.12	0.03
5	2	5	6	10	-1.41	0.16
6	2	1	2	2	0.58	0.56
7	0	0	0	1	-1.00	0.32

Table 8. Pedigree disequilibrium test analysis of D12S1056 alleles among ulcerative colitis families. * Only those alleles with a frequency > 5% are shown.

Haplotype* (D12S1056-D12S83)	Observed	Expected	Variance (O-E)	χ^2	<i>P</i>
3-7	4.26	4.73	0.96	0.22	0.64
4-7	2.30	2.55	0.46	0.14	0.70
5-7	1.94	3.50	0.73	3.33	0.07
4-9	4.31	2.95	0.70	2.65	0.10
5-9	2.09	1.57	0.36	0.77	0.38
3-10	1.05	1.72	0.71	0.63	0.43
3-11	3.06	1.99	0.30	3.76	0.05

Table 9. Haplotypes for the short tandem repeat markers D12S1056 and D12S83 in the indeterminate colitis population. * Only those haplotypes with a frequency > 5% are shown.

D12S1056 allele 3 and D12S83 allele 11 ($P=0.05$). These analyses are uncorrected for compounding of type 1 error.

Discussion

These data provide evidence of STR association within the IBD2 locus using family-based and population-based techniques. IBD2 disease-association was evident when using family-based analysis in the UC population. The STR D12S1056 showed disease association upon PDT analysis with over-transmission of allele 4 in UC families. Under-transmission of allele 3 was evident in the same family cohort. Given the robust nature of PDT analysis, this is an especially significant result.

Evidence of association between CD and IBD2 was less compelling than for UC. In case-control analyses, colonic CD (L2 and L3 groups) patients showed significant over-representation of D12S1022 genotype 3,4 compared to controls following correction of the global test for multiple analysis. Terminal ileal CD (L1 group) patients showed under-representation of D12S83 allele 8. The latter association appeared of uncertain significance, however, following correction of the global test for of allele homogeneity multiple analyses, with the global comparison between terminal ileal CD versus controls for D12S83 alleles only approaching the significance threshold of $P \leq 0.0125$, with estimation of the global exact test being $P=0.014$. It should be noted that the current study has low numbers of participants and thus lacks somewhat in statistical power. With increased numbers, and consequently increased power, it is not improbable that this global test could become significant.

Neither of these associations between D12S83 and CD subgroups could be confirmed using family-based studies of association, and thus may be due to the confounding effects of population substructure. This finding is however in agreement with previous studies, including Parkes et al. (123), who found that IBD2 appears to play a more significant role in UC susceptibility than in CD susceptibility.

One possible reason, however, for the lack of correlation between population-based and family-based testing may be that the CD family cohort has a rather heterogeneous nature. Owing to the small numbers of families for each CD subgroup, data from CD families were combined. An increased number of first-degree relatives of CD-affected individuals will therefore be required to absolutely confirm or refute the findings of disease association observed in the case-control association studies presented here. Such individuals are currently being actively accrued in order to address this problem.

In contrast to many earlier studies, evidence is presented for association of IBD2 with IC. Indeed, many groups have chosen to omit this population from genetic studies owing to the apparent diagnostic ambiguity that leads to a patient being classified with IC. There is strong evidence to suggest that IC should be considered a distinct clinical entity (99). IC patients have been shown to suffer from more fulminant disease, as well as have a higher incidence of complications following surgery. IC is considered as a separate and distinct IBD phenotype in the current study primarily based on clinical observations, long-term follow-up (average follow-up, 5.6 years), and the findings of a single specialist gastrointestinal pathologist's review. These data provide evidence that IBD2 is strongly associated with IC. D12S83 allele 11 is significantly over-transmitted

in IC families on likelihood ratio testing. Further strength to this argument is added by the finding of linkage disequilibrium between D12S1056 and D12S83 in the IC population based upon TDT analysis. This haplotype that displayed over-transmission included D12S83 allele 11, the same allele over-transmitted on likelihood ratio testing. Case-control studies partially confirmed this result with over-representation of allele 7 in the IC group. As with the case-control test discussed above, the global test of allele homogeneity only approached significance when using the Bonferroni method to correct for multiple testing ($P=0.027$). Therefore, the apparent allelic association should be viewed cautiously. As before, low study participant numbers does raise the possibility of type 2 statistical error, and it may therefore prove advantageous to repeat these analyses when further patients have been accrued.

Agreement between results from case-control studies, TDT, and other family-based tests in the IC population is important, since associations detected in case-control studies could be spurious due to the presence of population substructure. Given these data, it appears feasible to speculate that IC is genetically more similar to UC than to CD. Characterization of further genotype-phenotype relationships will lead to an increased understanding of the genetic basis of IBD. The majority of such relationships thus far have been described with the IBD1 susceptibility gene *NOD2*. Here, the presence of any of the three well-characterized disease-causing mutations has been shown to increase susceptibility to developing fibrostenosing ileocolic CD (31;33;35;36). There are however very few studies that have aimed to define genotype-phenotype relationships for IBD susceptibility loci. The majority of susceptibility loci studies have used patient groups that appear to be somewhat heterogeneous, in which variables such as disease

location and severity have not been adequately accounted for. In addition, few studies have considered IC as a separate IBD phenotype.

The overall aim of this study has been to improve IBD2 characterization by defining genotype-phenotype correlations. This in turn may aid the characterization of the IBD2 susceptibility gene(s). There essentially are two approaches to elucidating susceptibility genes: positional cloning and candidate gene approaches. Although positional cloning is a powerful means of locating a susceptibility gene, it suffers in that it is an extremely time consuming process. To date, there are no studies published relating to positional cloning attempts to define the IBD2 susceptibility gene. In contrast, candidate gene methods are relatively simple, but do suffer from a number of significant drawbacks, the most prominent of which is that there may be a large number of biologically plausible candidate genes within genomic region displaying a high degree of disease linkage (i.e. in close proximity to D12S83 in the case of IBD2). Many plausible IBD2 candidates have been identified (124), and genes studied to date include *AVIL*, which codes for the advillin protein and is involved in microvilli morphogenesis (126), and $\beta 7$ integrin, which is involved in enteric lymphocyte chemotaxis and retention of intraepithelial lymphocytes (127). None of these has as yet displayed convincing evidence of disease association, and the underlying cause of linkage/association within IBD2 remains unknown.

Future studies of IBD2 in this laboratory will follow a candidate gene approach, and will involve characterization of *IRAK-M*, a gene that is in close proximity to the peak region of linkage in IBD2. *IRAK-M* is a negative regulator of the Toll-like receptor

pathway, a key proinflammatory pathway of the innate immune system, and integral in the recognition of bacterial products such as lipopolysaccharide (58).

CHAPTER VII

PRELIMINARY ASSESSMENT OF INTERLEUKIN-1 RECEPTOR ASSOCIATED KINASE-M AND MYELOID DIFFERENTIATION FACTOR 88 AS INFLAMMATORY BOWEL DISEASE CANDIDATE GENES

Overview

BACKGROUND: The discovery of IBD susceptibility genes has proven extremely difficult, and to date only one such gene, *NOD2*, has been definitively implicated in disease pathogenesis. Definition of further susceptibility genes remains of vital importance in understanding the pathophysiological mechanisms underlying IBD susceptibility. The aim of this study is to determine whether further assessment of two biologically plausible IBD candidate genes, *IRAK-M* and *MyD88*, is worthwhile.

HYPOTHESES: 1) *IRAK-M* and *MyD88* are differentially expressed in IBD-affected colonic mucosa. This may be caused by mutation of regulatory elements within each gene. 2) *IRAK-M* and *MyD88* contain polymorphisms within coding, regulatory and adjacent genetic regions in a sizeable proportion of the study population. **METHODS:**

Candidate gene expression was measured from mRNA extracted from IBD-affected and disease free colonic mucosal samples using real-time PCR. Following quantification of gene expression, candidate gene coding and regulatory regions were screened for germline polymorphisms in a representative proportion of the entire study population. One-third of the screening cohort of 96 individuals had Crohn's disease, one-third had

ulcerative colitis, and one-third were population controls. PCR amplicons were designed to contain candidate gene coding and regulatory regions. Amplicons were then screened for polymorphisms using dHPLC. Polymorphic amplicons were sequenced to determine the nature of the variation. **RESULTS:** Both candidate genes were significantly over-expressed in IBD-affected colonic mucosa when compared to disease-free samples.

IRAK-M was found to contain four relatively commonly occurring variations. Two variations were within coding regions: a missense mutation within exon 5 and a synonymous single nucleotide polymorphism (SNP) within exon 12. SNPs were also discovered within introns 1 and 2, both of which were in close proximity to exon 2.

MyD88 contained three polymorphisms, all of which were non-coding. Two were located in the regulatory 3'-untranslated region, and one adjacent to exon 3, within intron 3.

CONCLUSIONS: Both *IRAK-M* and *MyD88* appear to be feasible candidate genes since they are differentially expressed in IBD-affected tissue and contain a number of commonly occurring polymorphisms. Polymorphisms defined in this study will eventually be characterized in the entire study population. The aim of the follow-up study will be to assess whether candidate gene polymorphisms identified here are associated with disease, and therefore increase susceptibility to IBD development.

Introduction

Identification of IBD susceptibility genes remains of paramount importance to understanding the mechanism of disease development. Many promising candidates have been studied, but only one gene, *NOD2*, has been conclusively proven to be involved in IBD pathogenesis.

Cytokine dysregulation is a prominent feature of all forms of IBD, with mucosal synthesis of pro-inflammatory cytokines, including interleukin (IL)-1 β , IL-6, IL-8, IL-16, and tumor necrosis factor-alpha (TNF α) being increased in both UC and CD. Despite of our poor understanding of the specific abnormalities leading to this proinflammatory cytokine excess, it is clear that inappropriate activation transcription factor nuclear factor κ B (NF- κ B) is a prominent feature of IBD (48). NF- κ B is involved in up-regulation of a variety of key regulators of the inflammatory response including pro-inflammatory cytokines, chemokines, inflammatory enzymes, adhesion molecules, and receptors. Activation of NF- κ B is normally induced by a variety of sources including cytokines and pathogens. However, the cause of NF- κ B over-activity in IBD is again, poorly comprehended. Inappropriate NF- κ B activation is likely a leading cause of the characteristic chronic gastrointestinal inflammation in IBD.

The Toll-like receptor (TLR) pathway is a receptor-mediated signaling cascade that is currently subject to a fair amount of scrutiny with respect to inappropriate NF- κ B activation in IBD. Toll-like receptors are a key component of the innate immune system and are involved in initiation of inflammatory responses through recognition of a wide variety of pathogens. Two promising IBD candidate genes within this pathway have been identified, each of which having already been discussed in great detail. The first is *IRAK-M*, a molecule that inhibits signaling through the TLR pathway (58). The hypothesis here is that disruption IRAK-M function could lead to uninhibited TLR stimulation in response to normal bowel luminal contents. Over a period of time, this would lead to the type of chronic inflammation characteristic of IBD. The second candidate is the adaptor protein MyD88. MyD88 serves as a link between TLRs and the serine threonine kinases

IRAK-1 and IRAK-4 (56). Activation of these kinase molecules is required to initiate downstream components of the signaling pathway leading to NF- κ B up-regulation. However, MyD88 has also been shown to interact with a number of other death-domain containing proteins other than IRAK-1/-4 following macrophage TLR stimulation. An example of such a protein is FADD, an intermediary in an apoptotic signaling cascade (71). To resolve inflammation, immune effector cells must be eliminated through apoptosis (73), and if this process were to somehow become disrupted (e.g. by mutation of MyD88), immune effector cells would persist and inflammation would continue for an abnormal period of time. It is therefore hypothesized that mutation of MyD88 leads to the type of chronic inflammation characteristic of IBD.

The aim of this study is to gather further evidence to prove that *IRAK-M* and *MyD88* are feasible IBD candidate genes. Compelling evidence has already been provided, at least for *IRAK-M*, to show that this gene may be involved in IBD pathogenesis. Data outlined in Chapters V and VI has shown that STR markers on chromosome 12q13, the genomic region containing *IRAK-M*, are associated with different forms of IBD. In this chapter, candidate gene colonic mucosal expression will be defined in both IBD-affected and normal tissue. Abnormal gene expression may well be indicative of mutation of a regulatory region within a candidate gene. Following this, candidate gene coding and regulatory regions will be screened using dHPLC. If a polymorphism is detected through this screening process, DNA sequencing will be used to fully characterize the candidate gene amplicon. This will allow sufficient information to be gathered for polymorphism genotyping in the entire population in future studies.

Materials and Methods

Gastrointestinal Mucosal Gene Expression

Methods for determining candidate gene colonic mucosal expression levels have been discussed in detail in Chapter IV. Briefly, colonic mucosal samples were obtained from two individuals with UC, two individuals with CD and two individuals with no history of IBD or colorectal, with the latter serving as controls. IBD-affected individuals in expression studies were not treated with immunosuppressive therapy prior to harvesting of mucosal samples. This is especially important given that immunosuppression may drastically alter gene expression patterns. mRNA was isolated from total RNA extracted from each mucosal sample. This mRNA was then used as a template for the cDNA synthesis, which was subsequently used in real-time PCR experiments. These experiments were performed in the following manner:

1) RNA Extraction

100mg of colonic mucosal tissue from UC, CD and normal mucosa was subjected to total RNA extraction following homogenization and cell lysis in guanidine isothiocyanate-containing buffer using an affinity column technique (RNeasy[®] Midi purification kit, Qiagen, Valencia, CA).

2) cDNA Synthesis

Collected cellular RNAs were then converted to cDNA using Invitrogen Superscript[®] First Strand Synthesis System for real-time RT-PCR (Invitrogen, Carlsbad, CA). 1µg of total cellular RNA was mixed with poly dT Primer 1µl (50µM oligo[dT]20) and 1µl of 10mM dNTP mix, and the total volume adjusted to 10µl with DEPC-treated

water. Reaction mixtures were then incubated at 65°C for 5 min and subsequently placed on ice for 1 min. The following were then added to mixtures: 2µl of 10x RT buffer, 4µl of 25mM MgCl₂, 2µl of 0.1M DTT, 1µl of RNaseOUT (40 U/l) and 1µl of SuperScript II RT (200U/l). Reactions were incubated for 50 min at 50°C. Termination of reactions was performed by incubating at 70°C for 15 min followed by chilling on ice. 1µl of RNaseH is then added and reactions were incubated for 20 min at 37°C.

3) Real-Time PCR Expression Analysis

Real-time PCR quantification was performed to compare candidate gene expression in diseased and disease-free colonic mucosal samples.

Primers were designed for the candidate genes *MyD88* and *IRAK-M* using Vector NTI software (107), and weighted to the 3' UTR. Primer sequences were as follows:

MyD88

5'AGCTGGCTCTGCTGGTCCTTCT3' and 5'TCCTGCATGGAACCAGTGGC3'

IRAK-M

5'TGTCCTCTCCATCTCCTCAGTGTG3' and 5'TTCCTACAACCCGTGCAAGCA3'

All reactions were performed on an ABI Prism 7700 Sequence Detection System (Applied Biosystems, Foster City, CA) using Brilliant® Quantitative PCR Kit (Stratagene, La Jolla, CA). *β-actin* was used as an internal standard (Proligo, La Jolla, CA). Primer sequences were as follows:

β-Actin

5'-AATCAAAGTCCTCGGCCACATTG-3'

and 5'-AAACTGGAACGGTGAAGGTGACAG-3'

Each reaction was performed in triplicate and contained 150nM of gene specific primers, 1x mastermix, and 1μl of cDNA (2nM final concentration) in a 25μl reaction. Reactions without cDNA were included as negative controls. Reaction conditions were as follows: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 20 sec, 1 min annealing at primer-specific temperatures (*IRAK-M* 54°C, *MyD88* 55°C, *β-actin* 54°C) and extension 72°C for 1 min. Signals were detected at annealing and extension phases.

Cycle threshold (Ct) values were obtained graphically for candidate genes as well as for the *β-actin* internal standard. Standard curves were constructed for all genes before calculation of gene-specific mRNA levels. Gene expression was normalized to *β-actin*. For each sample the mean of the three Ct values was calculated. Comparison of gene expression between control and disease samples was performed using a standard curve method, with gene expression in controls considered as normal. Candidate gene expression is given as a fold-change from this “normal” value.

PCR Amplification of Candidate Gene Coding and Regulatory Regions

PCR primers for *IRAK-M* and *MyD88* coding and regulatory were designed using the parameters discussed in Chapter IV. Primer sequences and a number of other characteristics of these amplicons are shown in Tables 10 and 11.

Table 10. PCR amplicons designed for IRAK-M using VectorNTI software. ^a Feature too large to be analyze as one fragment using dHPLC. ^b Exons 9 &10 analyzed as one amplicon due to their proximity.

Feature	Size (bp)	Position of Feature within Amplicon (bp)	PCR Annealing Temp (°C)	GC %	Upstream Primer	Downstream Primer
Promoter 1 ^a	537	1-537	67.0	45.3	TGTATGGAGGAGAGGCCGCC	CAGCTCACGGAGCCTCCACC
Promoter 2 ^a	770	1-663	67.0	62.6	AACGACACTCCCTCTCTACGAA	GTCGCAGCTGTCCAGAACAG
Exon 1	720	102-234	65.0	62.1	TGCCGTCGTGGAAGCAGGAT	CCGCGTTCATTGCAAACCGT
Exon 2	527	165-347	53.0	35.1	CAGAAGGCAGGTGAATATAT	AGGGAAGAAAGTGTCAACAAG
Exon 3	330	125-189	54.0	34.8	CATGGGAATGAACTTTAGA	ATAATGAGAGAACAAAGGCA
Exon 4	798	315-369	58.0	30.2	TGCATGTCACTGTTAGTTTT	GGAAAGTTGATATGGAAGAT
Exon 5	705	101-252	58.0	40.0	CTTCCACTGGACTCTGATTTG	CAGCCTCTTCTTCCTGTGCT
Exon 6	762	242-306	60.0	34.9	GTTAACATGGCGATACACTG	TCTCCAAGGACTCTCAGACT
Exon 7	275	101-215	55.0	41.1	GCCTATTCCCCATCCCACCTTAC	GCAGCTCTCCCAGGGATTCAAG
Exon 8	392	168-286	55.0	40.8	GAAACCCATCTGGAATTAAA	CACATACATACCAGGAGAAC
Exon 9 ^b	790	102-300	54.0	35.2	CGAATTAACCAGATATGAAG	GGAGGCACTTTCTTATCTAG
Exon 10 ^b		564-626				
Exon 11	420	123-287	62.0	41.2	GAGTAGTGTTAGATGATCCAA	TATGTTGGAATTCTCTGTTG
Exon 12.1 ^a	548	160-548	56.0	37.6	GGGCGTTAGCTAATCTTTAA	GGAGCTTCTGAAGAAGGATC
Exon 12.2 ^a	769	1-669	58.0	40.6	GGACAAAAAGCCAGAGAGCA	ACAAGCAGGCTAGTCAGGGA
3' UTR	642	33-468	58.0	39.9	GGGATGAATATGAACAGTACA	AGGATAGATCTTGGGAAGAA

Table 11. PCR amplicons designed for MyD88 using VectorNTI software. ^a Feature too large to be analyze as one fragment using dHPLC.

Feature	Size (bp)	Position of Feature within Amplicon (bp)	Optimized PCR Annealing Temp (°C)	GC %	Upstream Primer	Downstream Primer
Promoter 1 ^a	581	1-581	62.0	53.4	GGCTCTAATTCCTCTACCCA	GGGACGTCATGACACTAGGT
Promoter 2 ^a	811	1-700	63.0	63.7	GGCCATGTCTTCACCTGAGA	TCAAGAACAGAGACAGGCGG
Exon 1	536	109-436	63.0	67.4	CTTCTCGGAAAGCGAAAGCC	GCAGGAAATGGGGTCTCTCC
Exon 2	458	97-257	65.0	55.7	CCTTCCAGAAAGCCAGAACA	ACTTGCCCAAGGTCTCACAG
Exon 3	506	112-292	68.0	58.5	ATCTTGGGAAGGGTGCAGGG	AGACAACCACCACCATCCGG
Exon 4	494	101-192	64.0	55.7	CTGTCCAGCCTGGGCACAGT	GTCGCTTCTGATGGGCACCT
Exon 5	405	101-255	64.0	55.1	TAGATGGGGGATGGCTGTTG	GCTGCTGCAGACGTGTCTGT
3'-UTR 1 ^a	836	155-836	64.0	54.6	GTGCCCATCAGAAGCGACTG	GAAGGTTGGGCAGAAGCTGG
3'-UTR 2 ^a	678	1-678	62.0	51.6	ACTGATGCTTCAGTGCCTCT	GGGACACAAGAGAAAAGGAA
3'-UTR 3 ^a	816	1-649	62.0	46.3	CTCCCAGGAACAGCTAGGTG	GTATTCCTGGTTCTGCCTGC

Primers were custom-synthesized (Proligo, La Jolla, CA) and STRs amplified from genomic DNA using standard PCR apparatus (Perkin-Elmer Thermocycler, Norwalk, CT). PCR mixtures consisted of 100ng genomic DNA, 200nM of each dNTP, 2.5µl Thermophilic DNA 10x reaction buffer (Promega, Madison, WI), 1.5µM magnesium chloride, 1.0µM of both forward and reverse primers, and 2 units Taq DNA polymerase (Promega, Madison, WI) in a final 25 µl reaction volume. A touchdown PCR protocol was used to amplify candidate gene features, with thermal cycling parameters comprising an initial denaturation step of 95°C for 10 min followed by 2 cycles of denaturing at 94°C for 20 sec, annealing at 4°C above the primer-specific annealing temperature shown in Tables 10 or 11 for 60 sec and extension at 72°C for 45 sec. This was then followed by 2 cycles of denaturing at 94°C for 20 sec, annealing at the primer-specific temperature + 3.5°C for 60 sec and extension at 72°C for 45 sec. The primer annealing temperature was decreased in this manner until the cycle annealing temperature equaled the annealing temperature shown in Table 10 or 11. This was followed by 20 cycles of denaturing at 94°C for 20 sec, annealing at the primer-specific temperature for 60 sec and extension at 72°C for 45 sec, followed by a final extension phase at 72°C for 5 min. PCR products were screened for polymorphisms using dHPLC mutation analysis.

Polymorphism Discovery Protocol

The screening population consisted of 96 Caucasian individuals. Thirty-two were population controls, 32 had UC and 32 had CD. Of the CD population, 18 had colonic CD (Vienna classification groups L2/L3). Individuals with colonic CD were included in

SNP screening since IBD2 appears to increase susceptibility to developing colonic CD in this population.

Following PCR amplification, 5µl of the total reaction volume was taken from each individual sample and combined into 12 groups of 8 pooled samples. dHPLC mutation analysis was then performed at amplicon domain-specific heteroduplexing temperatures (see Table 12 below for heteroduplexing temperatures of polymorphic amplicons). If a polymorphism was evident in one pooled samples, the unpooled samples were analyzed individually to determine which reaction contained a polymorphic amplicon. Amplicon sequencing was then performed to determine the nature of the polymorphism.

PCR amplification of amplicons found to be polymorphic on dHPLC mutation analysis was performed prior to sequencing. PCR mixtures consisted of 150ng genomic DNA, 0.5mM dNTP mixture, 10µl 10x Cloned *Pfu* reaction buffer (Stratagene, La Jolla, CA), 1.5mM magnesium chloride, 0.5mM of the same forward and reverse primers used in screening experiments, and 5 units Cloned *Pfu* polymerase (Stratagene, La Jolla, CA), in a final 100µl reaction volume. Amplification was performed using the same amplicon-specific PCR protocols as outlined above. Amplicons were subsequently purified using the Wizard® PCR Preps DNA Purification System (Promega, Madison, WI). Following amplicon purification, their concentration was quantified spectrophotometrically (GeneQuant, Pharmacia, Dublin, OH). DNA fragments were sequenced using fluorescent cycle sequencing and analyzed on a DNA sequencer (model CEQ2000XL; Beckman-Coulter, Fullerton, CA).

Results

Gastrointestinal Mucosal Gene Expression

IRAK-M: Expression of *IRAK-M* was up regulated in both CD and UC colonic mucosa when compared to normal, un-inflamed mucosa from controls. Real-time PCR (RT-PCR) amplification illustrated *IRAK-M* to be expressed an average of 8 times more in both CD and UC colonic mucosa as opposed to un-inflamed tissue. This was based on RT-PCR analysis of two CD colonic mucosal samples, two UC colonic mucosal samples, and 2 “control” colonic mucosal samples.

MyD88: Expression of *MyD88* was up regulated in both CD and UC colonic mucosa when compared to uninflamed colonic mucosa. Real-time PCR amplification showed *MyD88* to be expressed an average of 3 times more in CD and 13 times more in UC colonic mucosa compared to uninflamed control tissue. These data are based on RT-PCR data from two CD colonic mucosal samples, two UC colonic mucosal samples, and 2 “controls” colonic mucosal samples.

dHPLC Mutation Analysis

Apparently polymorphic *IRAK-M* and *MyD88* amplicons using dHPLC mutation analysis are shown in Table 12. Also shown here are the temperatures and dHPLC buffer gradients at which heteroduplexing was evident, and the total number of amplicons from each group displaying polymorphisms.

Five candidate gene amplicons appeared to be polymorphic using dHPLC mutation screening. It should be noted however that the *MyD88* 3'-UTR3 amplicon appeared to contain polymorphisms within two different dHPLC domains (i.e.

Amplicon	Heteroduplexing Temperature (°C)	dHPLC Gradient Details			Number of Polymorphic Samples		
		Length of Gradient	% Buffer B at Start	% Buffer B at End	Con	UC	CD
<i>IRAK-M</i> Exon 2	52.1	2 min	64	68	2	3	5
<i>IRAK-M</i> Exon 5	53.6	2 min	65	65	2	3	6
<i>IRAK-M</i> Exon 12.2	53.2	2 min	64	68	3	6	5
<i>MyD88</i> Exon 3	62.9	2 min	62	66	5	6	13
<i>MyD88</i> 3' UTR3	57.2	2 min	62	72	7	8	17
	58.1	2 min	62	72	7	3	14

Table 12. Polymorphic candidate gene amplicons on dHPLC analysis of the screening population. Note that the *MyD88* 3'-UTR3 amplicon displayed different distributions of polymorphisms at two different heteroduplexing temperatures, implying that more than one polymorphism may exist within this amplicon. 'Buffer B' - 0.1M triethylammonium acetate solution containing 25% acetonitrile.

heteroduplexing was observed at two temperatures), with different individuals possessing polymorphic chromatograms at these two different temperatures. This strongly implies that this amplicon contains two distinct polymorphisms. Other amplicons did not display heteroduplexing on dHPLC analysis and were therefore non-polymorphic. These candidate gene regions were therefore excluded from further analysis.

Sequence Analysis of Polymorphic Amplicons

Following dHPLC candidate gene polymorphism screening, individual amplicons were sequenced to define the nature of polymorphisms. Seven polymorphisms were discovered. Sequencing chromatograms, the position of the mutation within the candidate gene amplicon and the effect of the polymorphism on the protein sequence (if any) are shown in Figures 7 – 11.

Figure 7. Polymorphisms within the 527bp *IRAK-M* exon 2 amplicon.

- A: The presence of two roughly equally sized peaks at the same position on the sequencing chromatogram denotes the presence of a polymorphism. This sequencing chromatogram has two peaks at position 81bp, one representing an adenosine and one representing a thymidine. This means that this individual is heterozygous for an A/T polymorphism at this position within the sequenced amplicon. Following sequence alignment with Vector NTI software, this was found to correspond to a previously described T→A polymorphism within intron 1 of *IRAK-M* (NCBI designation rs1185630). The position of this SNP within the 527bp exon 2 amplicon is shown in B. The rs1185630 SNP is 72bp upstream of exon 2.
- B: Diagrammatic representation of the 527bp exon 2 amplicon. The coding region is displayed as the orange arrow.
- C: A second polymorphism was discovered in the *IRAK-M* exon 2 amplicon. This individual is heterozygous a C/T polymorphism at position 387bp. This was shown to correspond to a previously described C→T SNP within intron 2 of *IRAK-M* (NCBI designation rs1882200) following sequence alignment. The position of this SNP in relation to exon 2 is shown in B. The rs1882200 SNP is 28bp downstream of exon 2.

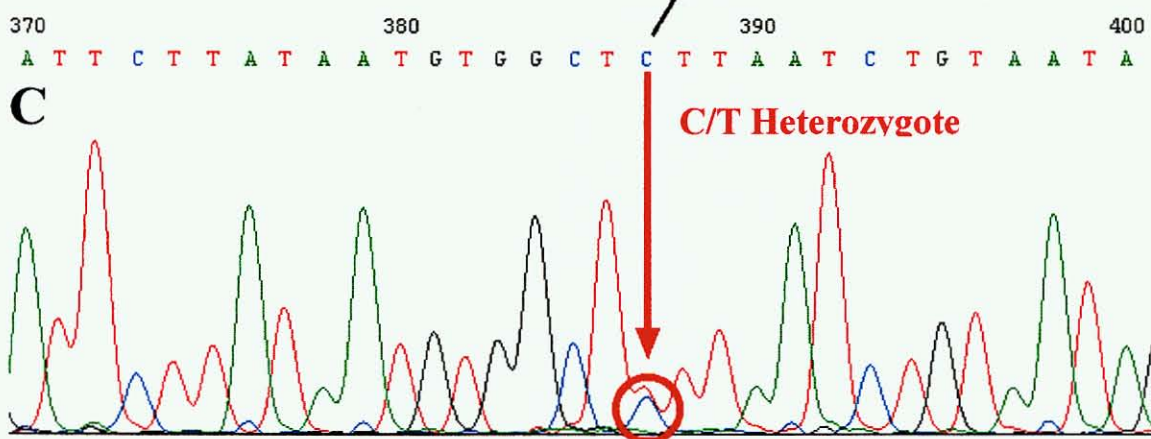
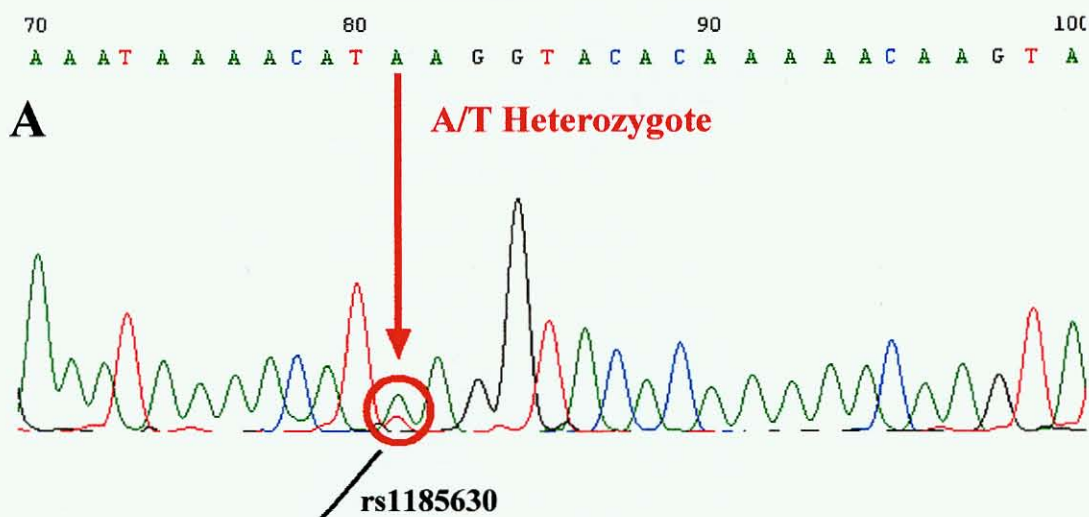


Figure 8. Polymorphism within the 705bp *IRAK-M* exon 5 amplicon.

- A: A polymorphism was discovered in the exon 5 amplicon. The individual shown is heterozygous for A/G polymorphism at position 65bp. Following sequence alignment using Vector NTI software, this was shown to correspond to the previously described 502G→A SNP within exon 5 of *IRAK-M* (NCBI designation rs1152888). The position of this SNP within exon 5 is shown in B.
- B: Diagrammatic representation of the 705bp exon 5 amplicon. The coding region is displayed as the orange arrow.
- C: Diagram of the IRAK-M protein. The 502G→A polymorphism causes a change in amino acid sequence of IRAK-M at position 147 in the amino acid chain. The wildtype residue at this position is a valine and the mutant is an isoleucine. This mutation therefore has the designation V147I.

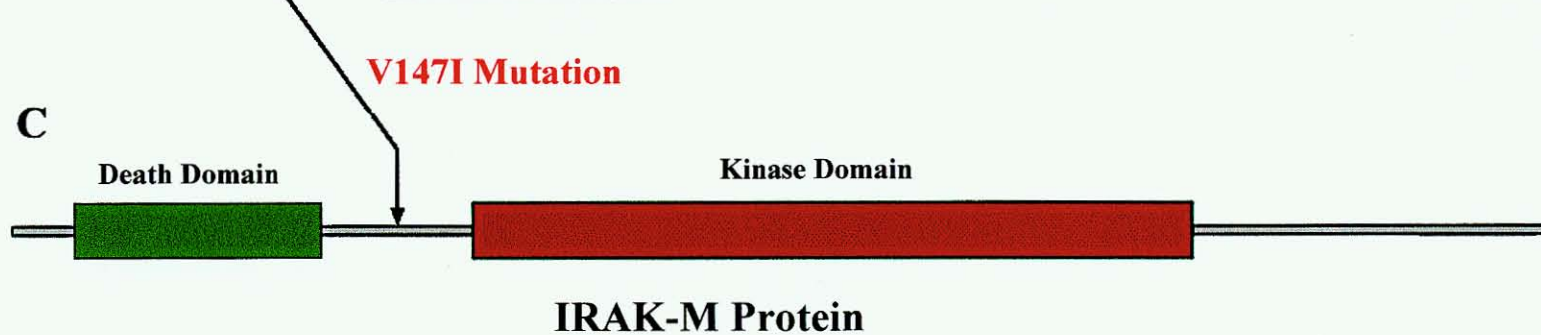
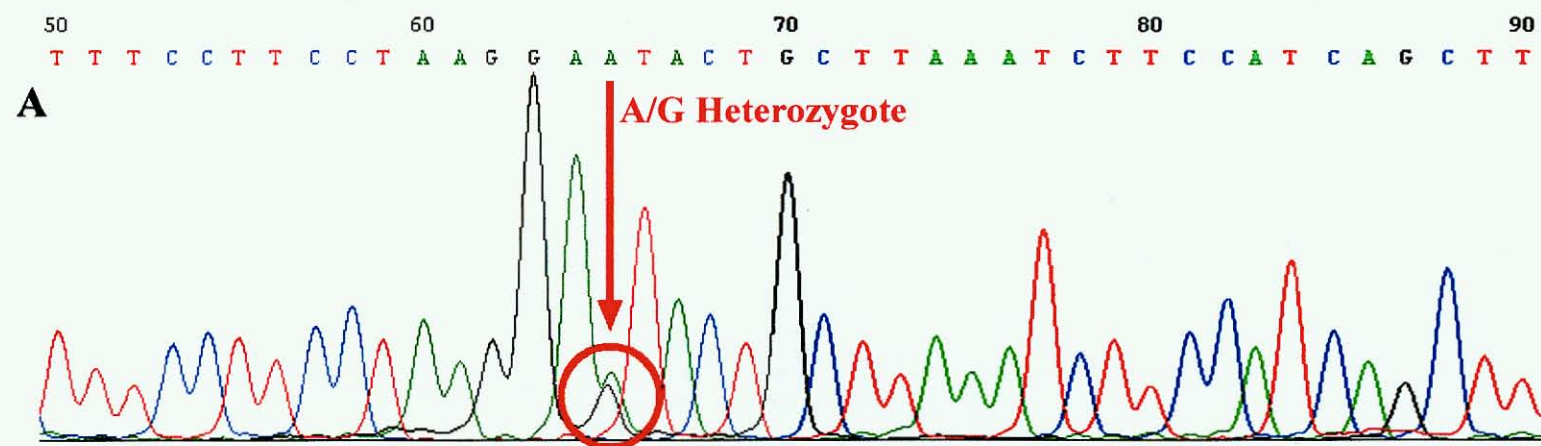


Figure 9. Polymorphism within the 769bp amplicon containing the 3'-end of *IRAK-M* exon 12 (exon 12.2).

A: A polymorphism was discovered in the exon 12.2 amplicon. The individual shown is heterozygous for an A/T polymorphism at position 101bp. Following sequence alignment using Vector NTI software, this within exon 12 of *IRAK-M*. Based on the position of the mutation within the *IRAK-M* mRNA, the polymorphism is given the designation 1752T→A. The 1752T→A SNP does not correspond to any previously described polymorphism and consequently has no NCBI designation. The position of the polymorphism within exon 12 is shown in B.

B: Diagrammatic representation of the 769bp exon 12.2 amplicon. The coding region is displayed as the orange arrow.

C: The 1752T→A polymorphism is a non-coding synonymous change and does not alter the amino acid sequence of *IRAK-M*. Both the wildtype T allele and the mutant A allele code for a proline at position 563 within *IRAK-M*. This mutation is therefore given the designation P563P.

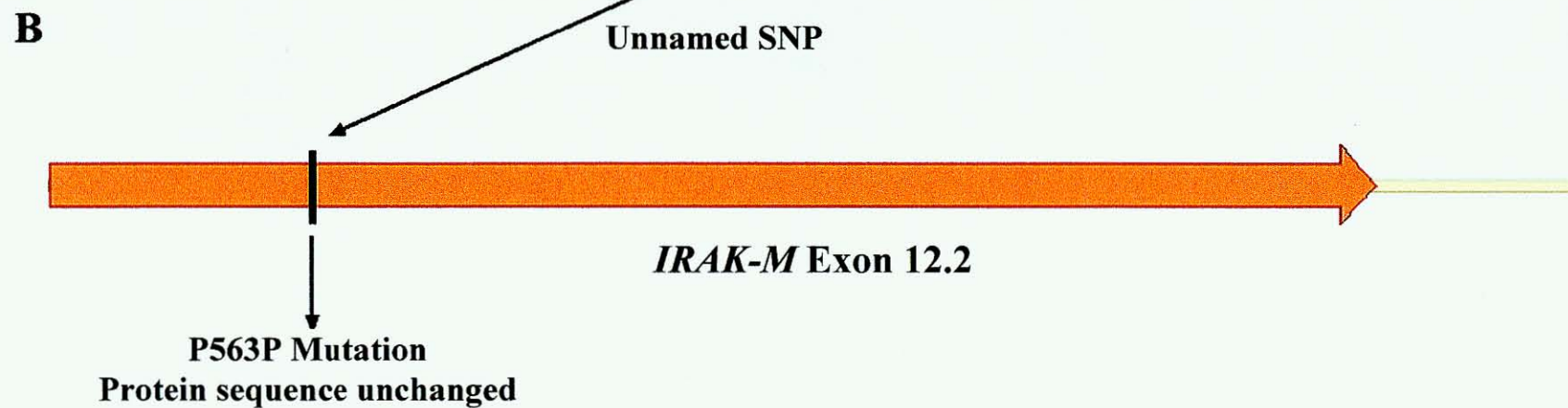
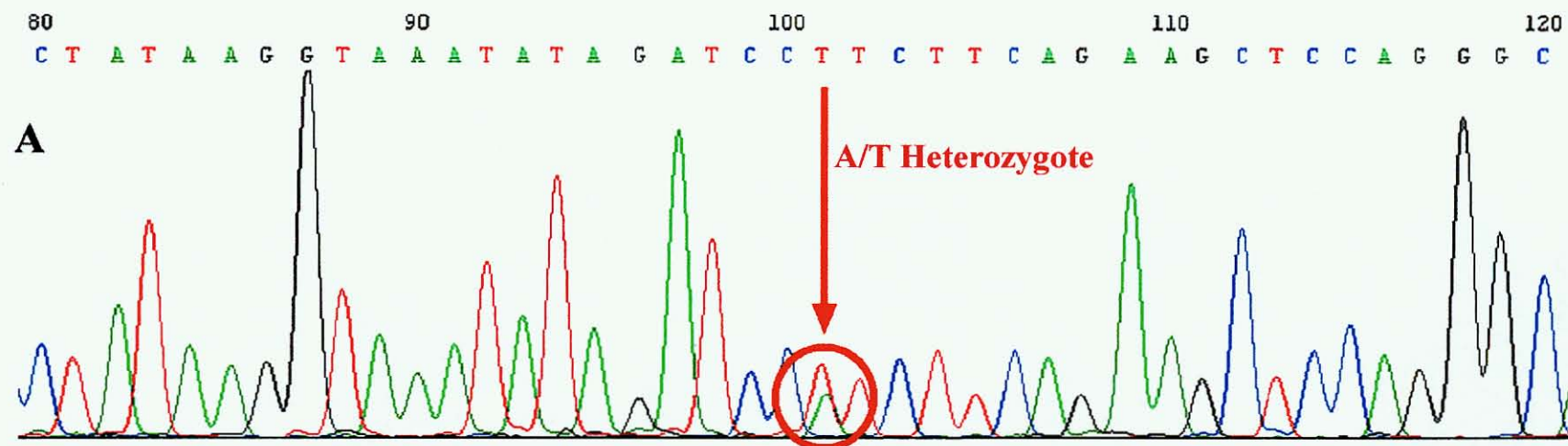


Figure 10. Polymorphism within the 705bp *MyD88* exon 3 amplicon.

A: A polymorphism was discovered in the exon 3 amplicon. The individual shown is heterozygous for C/G polymorphism at position 358bp. Following sequence alignment using Vector NTI software, this was shown to correspond to a previously described C→G SNP within intron 3 of *MyD88* (NCBI designation rs4988457). The position of this SNP in relation to exon 3 is shown in B. The rs4988457 SNP is 76bp downstream of exon 3.

B: Diagrammatic representation of the 705bp exon 3 amplicon. The coding region is displayed as the orange arrow.

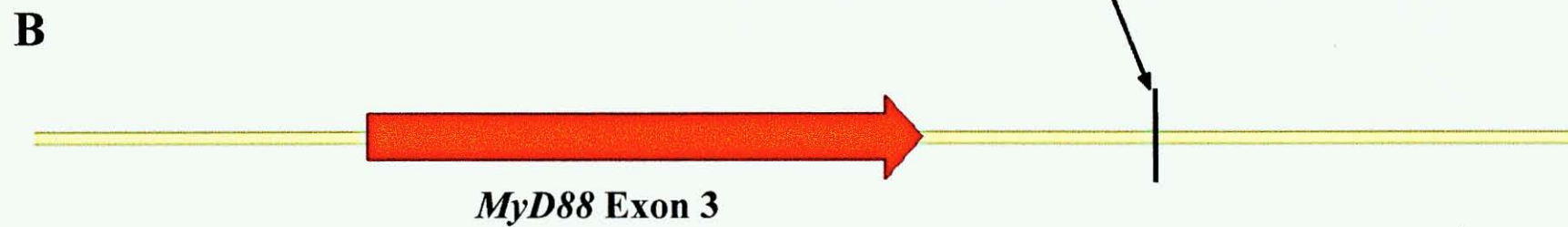
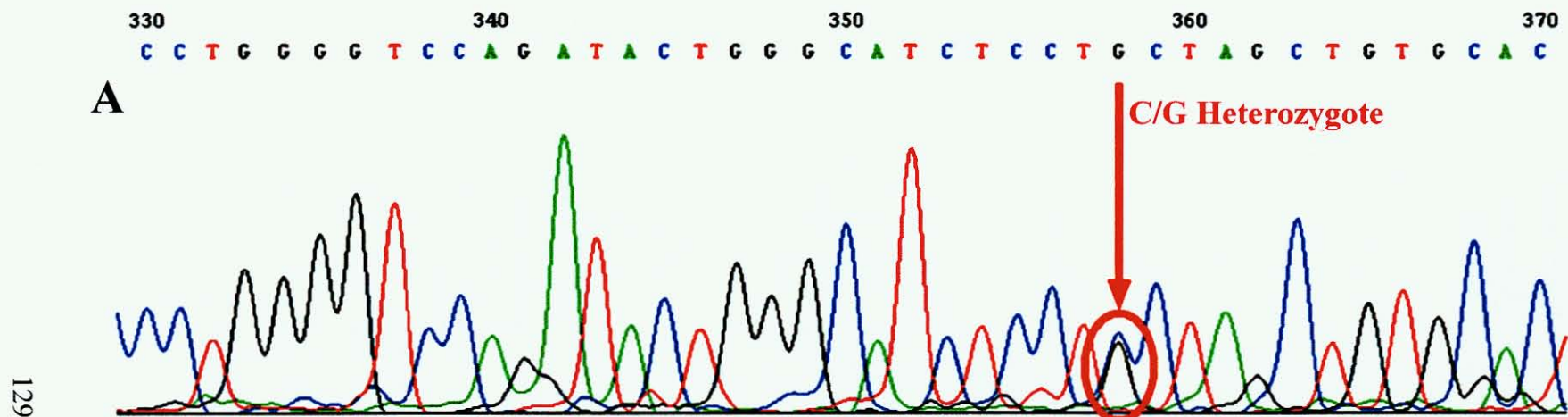
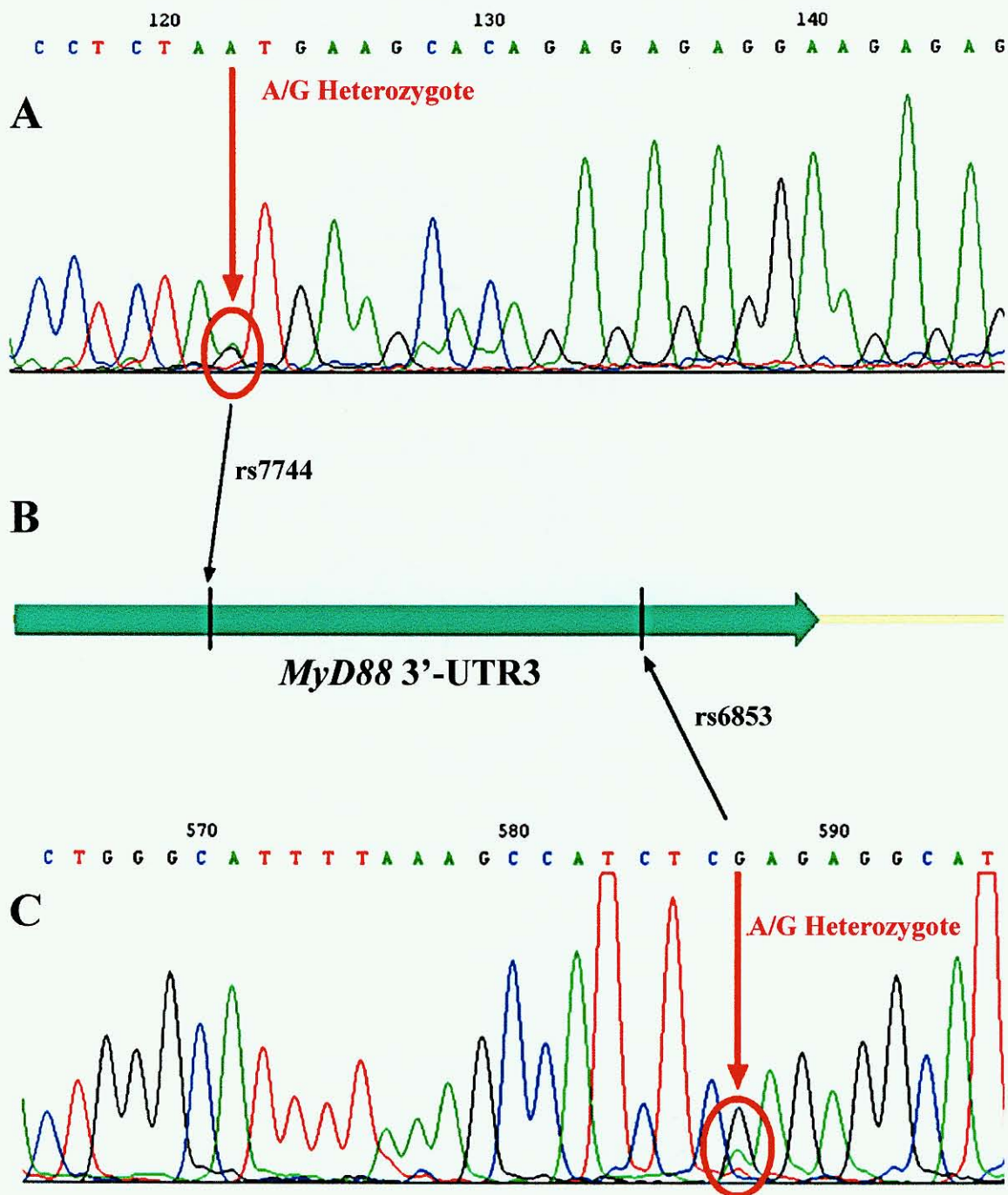


Figure 11. Polymorphisms within the 816bp amplicon containing the 3'-end of the 3'-untranslated region of *MyD88* (3'-UTR3).

A: A polymorphism was discovered in the 3' end of the 3'-UTR. The individual shown here is heterozygous an A/G polymorphism at position 122bp. Following sequence alignment, this was shown to correspond to a previously described A→G SNP within the 3'-UTR of *MyD88* (NCBI designation rs7744). The position of this SNP within the 3'-UTR is shown in B.

B: Diagrammatic representation of the 816bp amplicon containing the 3'-end of the 3'-UTR. The untranslated region is displayed as the green arrow.

C: A second polymorphism was discovered in the 3' end of the 3'-UTR. The individual shown here is heterozygous an A/G polymorphisms at position 587bp. Following sequence alignment, this was shown to correspond to a previously described A→G SNP within the 3'-UTR of *MyD88* (NCBI designation rs6853). The position of this SNP within the 3'-UTR is shown in B.



Discussion

Differential candidate gene expression has been identified in CD and UC colonic mucosal samples, with both *IRAK-M* and *MyD88* being significantly over-expressed in disease-affected colonic mucosal samples when compared to disease-free colonic mucosal samples. The magnitude of increased expression was significant in all IBD-affected tissues. For *MyD88* this ranged from a three-fold increase in expression in CD-affected colonic mucosa to thirteen times greater expression in UC-affected colonic mucosa. *IRAK-M* was over-expressed eight-fold in both CD- and UC affected colonic mucosa. The magnitude of this over-expression implies a significant degree of dysregulation of both candidate genes in IBD-affected colonic mucosa. There are likely two causes of this: an intrinsic genetic abnormality that leads to increased candidate expression (i.e. mutation of regulatory elements); or over-expression is a secondary phenomenon caused by dysregulation of another proinflammatory system. Obviously, at this stage there is insufficient evidence to implicate either as the cause of the candidate gene dysregulation in IBD. This question will be addressed through a process of elimination i.e. by screening genetic regions involved in regulation of candidate gene expression and characterizing polymorphisms within such regions.

Screening candidate gene amplicons for polymorphisms using dHPLC mutation analysis has proven an extremely powerful tool with seven SNPs being discovered using this methodology. All the polymorphisms discovered had a relatively high heterozygote frequency (i.e. heteroduplexing seen on chromatogram) in this small screening sample. Heterozygote frequency ranged from 10% (10/96 *IRAK-M* exon 2 samples showed heteroduplex formation, see Table 12) to 33% (32/96 *MyD88* 3'-UTR3 samples).

Therefore, all polymorphisms, at least at this stage of the study, appear to be relatively common.

The use of this screening protocol does raise a number of important criticisms. Firstly, rare polymorphisms may be overlooked when using a small screening population (i.e. 96 patients). Secondly, the composition of the screening population may well affect the outcome of screening, especially if a candidate gene polymorphism were to be involved in susceptibility to a very distinct phenotype. The first criticism is addressed when one considers the total study population. The screening population of 96 patients is derived from a total population of approximately 1000 subjects. The cohort used for dHPLC mutation screening therefore represents around 10% of the entire population. If a polymorphism was discovered, for example, in fewer than five individuals in the screening cohort, it is probable that the mutant allele frequency in the entire population would be extremely low. A mutant allele with such a low frequency is highly unlikely to have any major impact on disease susceptibility (at least in a cohort of 1000 patients) and further characterization of such a polymorphism would likely prove worthless. Attempts have been made to address the second criticism by including individuals with disease-associated IBD2 STR alleles in the screening population (see Chapter VI). The remainder of the screening population is made up of individuals with a broad array of different IBD manifestations. In this manner endeavors have been made to screen a population that is both tailored towards previous findings in this cohort, as well as being representative of the IBD population as a whole. In summary, dHPLC mutation analysis has proven to be an accurate and efficient means of detecting polymorphisms, and has significant

advantages over more cumbersome and costly techniques such as screening by sequencing.

A total of seven candidate gene polymorphisms were characterized using sequence analysis, four within *IRAK-M* and three within *MyD88*. Of the four polymorphisms in *IRAK-M*, two were within coding regions, one within exon 5 and one within exon 12. The 1752T→A polymorphism within exon 12 is a synonymous change and does not alter the amino acid sequence of the IRAK-M protein. Both the wildtype T allele and mutant A alleles code for a proline at position 563 within IRAK-M. Further consideration will therefore not be given to the P563P mutation since the 1752T→A polymorphism within *IRAK-M* exon 12 will have no effect on the functionality of the encoded IRAK-M protein.

In contrast, the 502G→A polymorphism within exon 5 (rs1152888) does cause a change in the IRAK-M protein, with the mutant A allele coding for an isoleucine instead of a valine at position 147. The structures of these two amino acids are shown in Figure 12. The position of the V147I mutation within IRAK-M is shown in Figure 8C. This mutation lies in the region connecting the kinase domain and the protein-protein interaction death-domain of IRAK-M. When the position of this is considered together with the nature of the mutation (i.e. hydrophobic valine → hydrophobic isoleucine), the likely functional effects of the change in amino acid sequence caused by the 502G→A SNP are likely minimal. However, IRAK-M is a protein that exists within the cytosol of macrophages and is therefore likely to be water-soluble. Hydrophobic amino acids in regions connecting the two functional motifs, such as the residue found at position 147, are likely to be folded towards the center of the molecule, leaving more hydrophilic

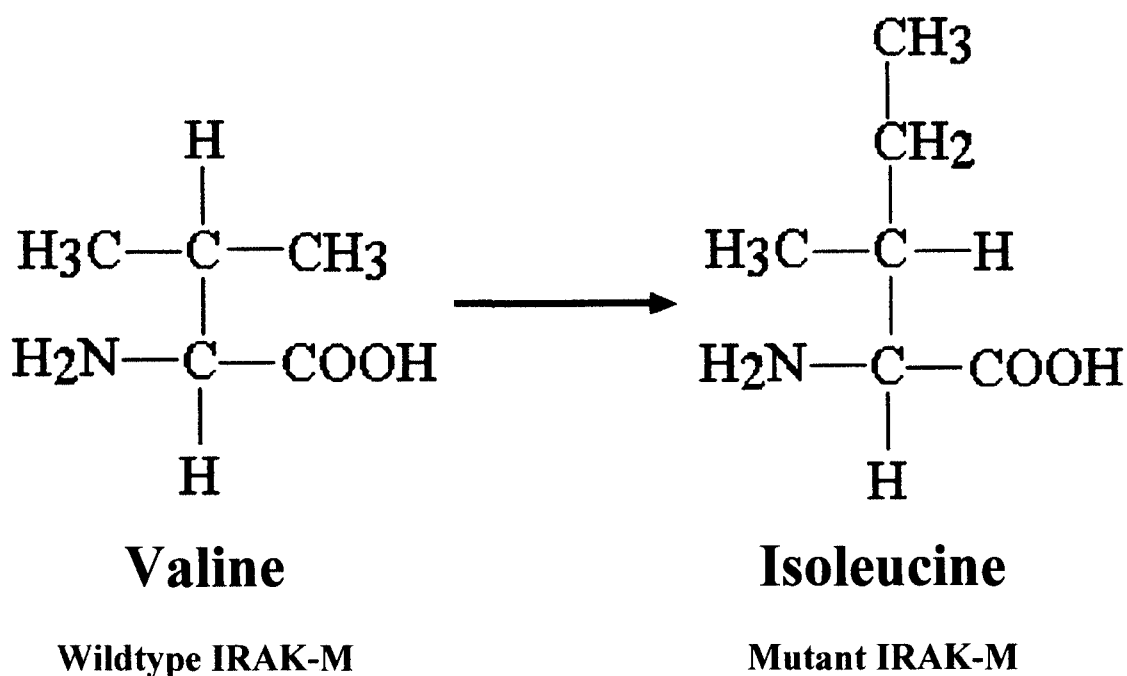


Figure 12. Molecular structure of valine and isoleucine. A valine residue is found at position 147 of the wildtype IRAK-M and an isoleucine is found in the same position in mutant IRAK-M. Both valine and isoleucine are classed as hydrophobic amino acids.

residues exposed to the water-rich cytosol. A mutation at such a position could alter the shape of the protein and thus modify its functionality. As illustrated in Figure 12, isoleucine contains more methyl side chains than valine. The increased ‘bulkiness’ of isoleucine in mutant IRAK-M may therefore alter the conformation of the protein. This in turn may affect its interaction with other signaling molecules, its ability to be activated and so on. Whether this mutation appears to predispose individuals to developing IBD will be assessed in Chapter VIII.

The two SNPs identified within the 3'-UTR of *MyD88* (rs7744 and rs6853) also appear to be of great interest. The 3'-UTR is an integral component of an mRNA, and is heavily involved in regulation of gene expression. These regulatory functions appear to be both diverse and numerous, and include control of nuclear export, polyadenylation status, subcellular targeting, and rates of translation and degradation of mRNA (128). It is very possible that mutation of a 3'-UTR may affect gene expression. Indeed, a number of human diseases have been linked to disruption of 3'-UTR mediated events. An example of such a disease is myotonic dystrophy, an autosomal-dominant multi-system disorder that has an incidence of about 1:8500. This disease is caused by an expanded number of trinucleotide (CTG) repeats in the 3'-UTR of a cAMP-dependent protein kinase gene (128). Other examples of “3'-UTR-mediated diseases” include α -thalassemia, Fukuyama-type congenital muscular dystrophy, and possibly acute myelogenous leukemia (128). The mutations within the 3'-UTR of *MyD88* may also influence disease pathogenesis by altering gene expression. Differential expression of *MyD88* has been identified in IBD-affected colonic mucosa. Therefore, the distribution of mutant rs7744 and rs6853 SNP alleles in the entire study cohort will be of great interest.

Three intronic mutations were also discovered. Two were within *IRAK-M*: one was 72bp upstream of exon 2 in intron 1 (rs1185630), and the other was 28bp downstream of exon 2 in intron 2 (rs1882200). The final SNP (rs4988457) was in intron 3 of *MyD88*, being located 76bp downstream of exon 3. Although intronic SNPs do not appear to be as ‘exciting’ as coding SNPs at first glance, there is growing evidence to suggest that like the 3’-UTR, introns play an important role in the regulating gene expression (129;130). Introns, and the act of their removal by the spliceosome, can influence many stages of the mRNA metabolism, including initial rates of gene transcription, pre-mRNA editing and polyadenylation, and nuclear export, translation and decay of the mRNA (130). These functions are dependant on the nucleotide sequence of the intron, and mutation of intronic sequences could alter gene expression, which in turn may lead to disease susceptibility. It is therefore possible that the three candidate gene polymorphisms residing within introns may be involved in IBD susceptibility.

In summary, initial assessment of *IRAK-M* and *MyD88* show that they both appear to be reasonable IBD susceptibility gene candidates. Both are over-expressed in IBD-affected colonic mucosa. Furthermore, both genes contain commonly occurring mutations within coding regions or gene elements that could be involved in regulation of expression. The next phase of this study will be to characterize candidate gene polymorphisms in the entire study population. The aim here will be to define SNP associations in specific IBD sub-phenotypes. Six of the seven discovered SNPs will be characterized. The seventh SNP, *IRAK-M* 1752T→A will not be characterized. As mentioned previously, this polymorphism is a synonymous change within *IRAK-M* exon 12, and consequently will have no effect on *IRAK-M* function.

CHAPTER VIII

IS THERE A ROLE FOR *IRAK-M* AND *MYD88* IN SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE?

Overview

BACKGROUND: Six SNPs have been identified within the two candidate genes using dHPLC mutation analysis and sequencing. Three of these lie within the gene encoding the anti-inflammatory inhibitor of TLR pathway signaling, *IRAK-M*, and three lie within the gene encoding *MyD88*, an adaptor protein shown to be an intermediary in a macrophage apoptotic pathway and therefore presumably involved in limitation of inflammatory responses. **HYPOTHESIS:** Mutation(s) of the genes encoding *IRAK-M* and *MyD88* cause abnormal functioning of their respective gene products, which in turn, leads to the type of chronic inflammation characteristic of IBD. **METHODS:** The study population consisted of 422 Caucasian IBD patients, 101 population controls with no history of IBD or colorectal cancer and 275 first-degree relatives of IBD patients. The six candidate gene polymorphisms were genotyped using SNP-specific PCR. Mutant allele frequencies were examined initially in unstratified IBD populations, and then in groups stratified on the basis of disease characteristics. Population- and family-based methods were used to analyze candidate gene SNP data. **RESULTS:** Preliminary evidence is provided to suggest that *MyD88* may be involved in susceptibility to late onset CD in patients with a family history of IBD. The frequency of the 3'-UTR rs7744 mutant allele

was significantly higher in late onset CD than in both early onset CD (mutant allele frequencies 30% and 15% respectively; $P=0.005$) and controls (mutant allele frequency: 13%; $P=0.003$). Furthermore, the 3'-UTR rs6853 SNP was significantly more frequent in CD patients with a family history of IBD than in CD patients with no family history of IBD (mutant allele frequency: 15% vs. 7%; $P=0.015$). Both results were significant following correction for multiple analyses. In contrast to *MyD88* SNPs, no evidence of *IRAK-M* SNP association was found in this IBD population. **CONCLUSIONS:** These results provide preliminary evidence to suggest that *MyD88* may be involved in susceptibility to late onset CD in patients with a family history of IBD. This finding is however viewed with some caution given the relatively small IBD population size and the lack of corroboration of the positive findings from population-based methods by additional statistical methodology. *IRAK-M* however plays no role in IBD susceptibility in this population

Introduction

The rationale for further study of the two IBD candidate genes has been discussed in depth in the previous chapter. Both *IRAK-M* and *MyD88* contain commonly occurring mutations within coding regions and/or gene elements with potential regulatory effects. Furthermore, each candidate is over-expressed in IBD-affected colonic mucosa. For these reasons, *IRAK-M* and *MyD88* are believed to be reasonable IBD susceptibility gene candidates. The polymorphisms identified in the previous chapter therefore require characterization in the entire study population.

The aim of this study is to determine whether the candidate gene SNPs identified in Chapter VII are associated with IBD. This will be assessed by performing population- and family-based testing in well-defined phenotypic groups. Phenotypic characteristics studied will include age of disease onset, presence or absence of extra-intestinal disease manifestations and family history of IBD. Furthermore, the CD population will be subdivided both on the basis of the Vienna Classification (98), as well on a number of other disease behavior characteristics (presence of perianal disease, fistulae not involving the perianal region and fibrostenosing behavior patterns). The importance of analysis at the sub-phenotype level has been proven by studies of the IBD susceptibility gene *NOD2*, where the degree of association is much stronger in CD subgroups than in the entire CD population (31;35;36). Increasing subgroup clinical homogeneity therefore seems to increase the strength of association, at least in the instance of *NOD2*. It is for this reason that candidate gene SNPs will only be characterized in Caucasians, given that this ethnic group comprises over 95% of the entire IBD population.

Candidate gene SNPs will be genotyped in individuals with IBD, as well as in population and family controls. The frequency of mutant SNP alleles will then be compared between disease and control populations (or between disease populations in the instance of the sub-phenotypic groups). A positive outcome would be a significantly increased frequency of one or more mutant SNP alleles in disease-affected populations when compared to controls. This would mean that the SNP is associated with IBD and the candidate gene is likely involved in IBD susceptibility.

Materials and Methods

Patients and Families

The University of Louisville Institutional Review Board approved this study. Written informed consent was obtained from all subjects. Patients were derived from the same population as described in previous chapters. The population-based association study cohort consisted of a total of 523 Caucasian individuals, including 208 unrelated individuals with CD (61% women), 159 with UC (53% women), 55 with IC (68% women), and 101 population controls (65% women). Demographic data for this population are shown in Table 13. For the purposes of more accurate phenotyping, the CD population was subdivided based on the Vienna classification (98). The CD population was also divided in an alternative manner, in order to include the following as distinct phenotypes: perianal fistulae, fistulae not involving the perianal region and gastrointestinal obstruction with or without the presence of fistulae. CD phenotype data are shown in Table 14.

Two hundred seventy-five non-affected first-degree relatives of Caucasian IBD patients were available for this study. There were 64 CD families (28 triads, 30 discordant sibling pairs [DSPs]), 6 triad/DSP combinations), 38 UC families (8 triads, 22 DSPs, 8 triad/DSP combinations), and 14 IC families (4 triads, 6 DSPs, 4 triad/DSP combinations). In addition, there were 16 CD families, 14 UC families, and 9 IC families who could not be classified as either a triad or DSP. The UC and CD family cohorts were subdivided on the basis of the same proband phenotypic characteristics, as described above. Subgroup analysis was not performed in the IC family cohort owing to the low frequency of these families in the study (n=14).

	Crohn's Disease n=208	Ulcerative Colitis n=159	Indeterminate Colitis n=55
Average Length of Clinical Follow-Up (Range) (yrs)	4.6 (0-16)	4.3 (0-14)	5.6 (0-12)
Age of IBD Onset (Range) (yrs)	29.2 (6-81)	34.4 (10-75)	29.7 (12-64)
Surgical Treatment for IBD Complications	75%	67%	82%
First-Degree Relative with IBD	36%	20%	22%
Extra-intestinal Manifestations ^a	43%	36%	49%

Table 13. Demographic data for IBD patients in the candidate gene study.

^a Extra-intestinal manifestations are defined as IBD-related disease outside the gastrointestinal tract involving joints, eyes, hepatobiliary system or the skin.

	<i>N</i>	%
<u>Age of Diagnosis</u>		
< 40 years (A1)	133	64%
> 40 years (A2)	36	17%
Unknown	39	19%
<u>Disease Behavior</u>		
Non-stricturing, non-penetrating (B1)	86	41%
Stricturing without penetration (B2)	19	9%
Penetrating (B3)	103	50%
<u>Disease Location</u>		
Terminal ileum (L1)	58	29%
Colon (L2)	49	24%
Ileocolon (L3)	59	29%
Upper gastrointestinal tract (L4)	7	3%
Unknown	30	15%
<u>Alternative Classification of Disease Characteristics</u>		
Perianal disease	52	25%
Fistula not involving perianal region	69	33%
Gastrointestinal stricture	46	22%

Table 14. Disease characteristics of the CD population used in candidate gene studies.

Patients are subdivided using the Vienna Classification (98), as well as a number of other CD characteristics.

The initial diagnosis of IBD was determined through conventional radiological, endoscopic, and/or histopathologic findings. Histology was available in all cases. Following the initial diagnosis, a single gastrointestinal pathologist with a particular interest in IBD reviewed all histology. Previous studies have shown that interobserver variation in histological diagnosis can be a significant confounding variable (97). If there was disagreement between the initial diagnosis and that of the specialist pathologist, the latter diagnosis was used for purposes of this study. The approach used here is likely to maximize diagnostic consistency, thus increasing the clinical homogeneity of IBD subgroups.

Characterization of Candidate Gene Polymorphisms

IRAK-M and *MyD88* SNP-specific PCR primers and fluorogenic probes were designed as per the specifications given in Chapter IV. The sequences of primers and probes are shown in Table 15. Reaction mixtures consisted of 300nM of each oligonucleotide primer, 100nM fluorogenic probes, 50ng template DNA, and 2x TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA) in a total volume of 11 μ l. The amplification reactions were performed on a Perkin-Elmer Thermocycler (Norwalk, CT) with two initial hold steps (50°C for 2 min, followed by 95°C for 10 min) and 40 cycles of a two-step PCR (92°C for 15 sec, 60°C for 1 min). The fluorescence intensity of each sample was measured post-PCR in an ABI Prism 7900HT sequence detection system (Applied Biosystems, Foster City, CA), and candidate gene SNP genotypes were determined by the fluorescence ratio of the nucleotide-specific fluorogenic probes.

SNP	Oligonucleotide	Oligonucleotide Sequence
<i>IRAK-M</i> Intron 1 T→A rs1185630	Forward Primer	5'-ACATTAGCAGAAGGCAGGTGAATAT-3'
	Reverse Primer	5'-TCTCTGCTGTGGAAATTGCAA-3'
	Wildtype Probe	FAM-AATAAAACATTAGGTACACAAA
	Mutant Probe	VIC-AATAAAACATAAGGTACACAAA
<i>IRAK-M</i> Intron 2 C→T rs1882200	Forward Primer	5'-TCGAGCTATTCATTTAATTACAAACTATGGT3'
	Reverse Primer	5'-AATGTCTTGGATAAAATGCAATATTCTG-3'
	Wildtype Probe	FAM-CAGATTAAGAGCCACATTA
	Mutant Probe	VIC-TTACAGATTAAAAGCCACATTA
<i>IRAK-M</i> Exon 5 502G→A rs1152888	Forward Primer	5'-GGTTATAATGGTGATGTTTCCTTCTTAGC-3'
	Reverse Primer	5'-TCTCTCCTTCTCCAATTAGGAAGTCT-3'
	Wildtype Probe	VIC-CCTTCCTAAGGAGTACTGCT
	Mutant Probe	FAM-CCTTCCTAAGGAATACTGCTT
<i>MyD88</i> Intron 3 C→G rs4988457	Forward Primer	5'-CATCGAAAAGAGGTTGGCTAGAA-3'
	Reverse Primer	5'-TTCAGGAAGGGCCCACTGT-3'
	Wildtype Probe	FAM-CATCTCCTCCTAGCTGT
	Mutant Probe	VIC-CATCTCCTGCTAGCTG
<i>MyD88</i> 3'-UTR A→G rs7744	Forward Primer	5'-CTGGAAAGGACCCAATGTACCA-3'
	Reverse Primer	5'-TGTTGTGTGAGTTTAAGCAGCTCTCT-3'
	Wildtype Probe	FAM-TTTATACCTCTAATGAAGCACAG
	Mutant Probe	VIC-TACCTCTAGTGAAGCACA
<i>MyD88</i> 3'-UTR A→G rs6853	Forward Primer	5'-GCACCTGTCCCCCTTTAATACTG-3'
	Reverse Primer	5'-TTGAAATTATTTTAATGCGTACAAAACATG-3'
	Wildtype Probe	FAM-CATCTCAAGAGGCATC
	Mutant Probe	VIC-CCATCTCGAGAGGCA

Table 15. Primers and fluorescently labeled oligonucleotide probes used in candidate gene SNP-specific PCR genotyping experiments. The position of the SNP within each probe is marked in red. FAM: 5-(6)-carboxyfluorescein. VIC®: a proprietary fluorescent dye produced by Applied Biosystems (Foster City, CA).

Statistical Analysis

Statistical analysis of candidate gene SNP data has been discussed at length in Chapter IV. Briefly, SNP data will be analyzed in two manners. First, candidate gene polymorphism frequency data will be analyzed in order to attempt to define associations between SNP alleles and IBD. The primary approach will be to perform case-control analyses to examine allele frequencies in CD, UC, IC, and in the IBD population when considered as a whole. Case-control will be augmented by estimation of candidate gene haplotype frequencies and family-based approaches. Correction for compounding of type 1 error will be performed using the Bonferoni method.

Following case-control analyses, attempts will be made to define genotype-phenotype correlations for candidate gene SNPs. This may be a powerful means of detecting association given the relatively small size of the study cohort. That is, stratifying the patient population into more clinically homogenous sub-groups may increase the degree of SNP association. Again, population-based, haplotype and family-based tests to study candidate gene genotype-phenotype correlations will be used. Correction for multiple analyses will be performed in a similar manner as described above.

Results

Candidate gene SNP genotype and allele frequencies and P-values for comparison of allele frequencies using case-control methods are shown in Table 16. The overall frequency of each candidate gene mutant SNP allele varied widely in the study

population, with the *IRAK-M* intron 2 rs1882200 mutant allele being most common (28% IBD patients, 26% controls) and the *MyD88* intron 3 rs4988457 mutant allele being least common (6% IBD patients, 7% controls). Population-based approaches were initially to compare candidate gene mutant allele disease frequencies between the IBD, UC, CD, and IC groups and the control population. As shown in Table 16, no associations were evident between disease populations and controls for the six candidate gene SNPs using population-based methods.

Following SNP analysis using population-based approaches, haplotype and family-based analyses were performed in IBD, UC, CD, and IC groups. Haplotype frequencies were estimated from SNP allele data using an expectation maximization method (see Chapter IV for a detailed description of these methods). These analyses did not yield any evidence of SNP association. Similarly, family-based approaches (PDT and likelihood ratio testing) did not reveal any evidence of association.

Analysis of Genotype-Phenotype Relationships

Population-Based Association Tests: Frequencies of candidate gene mutant alleles were calculated for the disease characteristics listed in Tables 13 and 14 in CD and UC populations. This was not performed in the IC population owing the low numbers of individuals in this population (n=55). Population-based analyses at the sub-phenotype level using a cohort of this size would be meaningless. Mutant allele frequencies for the *IRAK-M* and *MyD88* sub-phenotypes described in Table 13 are shown in Tables 17 and 18 respectively.

Table 16. *IRAK-M* and *MyD88* genotype and allele frequencies for patients and controls.

-/- Homozygous wildtype; +/- Heterozygous; +/+ Homozygous mutant

^a P-value for comparison of allele frequencies between controls and patients.

SNP	Cohort	Genotype			Mutant Allele Frequency	<i>P</i> ^a
		-/-	+/-	+/+		
<i>IRAK-M</i> Exon 5 rs1152888	IBD	334	55	5	8%	0.164
	CD	166	28	1	8%	0.168
	UC	120	19	1	8%	0.181
	IC	48	8	3	12%	0.934
	Control	81	15	4	12%	
<i>IRAK-M</i> Intron 1 rs1185630	IBD	227	65	2	12%	0.257
	CD	109	26	1	10%	0.262
	UC	81	31	1	15%	0.909
	IC	37	8	0	9%	0.239
	Control	57	16	3	14%	
<i>IRAK-M</i> Intron 2 rs1882200	IBD	170	145	23	28%	0.692
	CD	87	65	11	27%	0.987
	UC	61	55	10	30%	0.519
	IC	22	25	2	30%	0.667
	Control	38	39	1	26%	
<i>MyD88</i> Intron 3 rs4988457	IBD	285	36	1	6%	0.575
	CD	139	15	1	5%	0.617
	UC	105	15	0	6%	0.888
	IC	41	6	0	6%	0.977
	Control	67	9	1	7%	
<i>MyD88</i> 3'-UTR rs7744	IBD	221	104	5	17%	0.410
	CD	99	52	4	19%	0.225
	UC	85	42	1	17%	0.534
	IC	37	10	0	11%	0.508
	Control	58	21	1	14%	
<i>MyD88</i> 3'-UTR rs6853	IBD	250	61	4	11%	0.370
	CD	122	28	1	10%	0.651
	UC	90	28	2	13%	0.158
	IC	38	5	1	8%	0.827
	Control	63	10	1	8%	

Allele frequencies were compared using Fishers Exact test. One significant difference was evident for *MyD88*: the frequency of the 3'-UTR rs6853 A→G mutant allele was significantly more common in individuals with a family history of IBD when compared to those with no history of IBD (22/144 alleles [15%] vs. 18/246 alleles [7%] respectively; $P=0.015$). This was significant following at the Bonferroni-corrected significance level of $P\leq 0.0167$. This significance level was used since three comparisons are made to test the null hypothesis "there is no difference in *MyD88* SNP mutant allele frequency when considering presence or absence of family history of IBD in CD" (i.e. 3 SNPs, 2 comparison groups, 1 disease group, $\therefore P=0.05/3$). The rs6853 mutant allele frequency did not differ when comparing the family history positive CD group to the control population (rs6853 frequency in CD with family history of IBD: 22/144 vs. control frequency: 14/168; $P=0.07$). This result should therefore be viewed with some caution. In contrast, none of the *IRAK-M* mutant SNP alleles appeared to be associated with any of the phenotypic variables outlined in Table 13.

Following this, the additional CD phenotypic characteristics described in Table 14 were considered. Mutant allele frequencies were calculated for each sub-phenotype (allele frequencies not shown) and comparisons were made in the same pair-wise fashion as described above. The *MyD88* 3'-UTR A→G SNP (rs7744) mutant allele was significantly more frequent in the A2 Vienna group (age of disease onset ≥ 40 yr) than in the A1 group (age of onset < 40 years) group (21/70 alleles [30%] vs. 37/250 alleles [15%] respectively; $P=0.005$). This was significant at the Bonferroni-corrected level of $P\leq 0.0167$. Furthermore, the rs7744 mutant allele in the late onset CD group was also significantly more frequent than in the control population (rs7744 frequency in late onset

Table 17. *IRAK-M* mutant allele frequencies for CD and UC populations at the sub-phenotype level. Comparisons of mutant allele frequencies were performed between different subgroups of the same phenotypic variable (i.e. CD patients with a family history of IBD vs. CD patients with no family history of IBD) using Fishers Exact test. Allele frequency is given as: “mutant allele frequency / total number of alleles”. None of these comparisons were statistically significant.

^a Early age of onset in this instance was defined as ≤ 20 years for CD and ≤ 40 for UC.

The frequency of UC with onset ≤ 20 years was low (13% of total UC population) and comparisons would have been meaningless. Comparisons for CD with onset ≤ 40 years are described below (see Vienna Classification).

Crohn's Disease							Ulcerative Colitis					
<i>IRAK-M</i>		<i>IRAK-M</i>		<i>IRAK-M</i>			<i>IRAK-M</i>		<i>IRAK-M</i>		<i>IRAK-M</i>	
Intron 1		Intron 2		Exon 5			Intron 1		Intron 2		Exon 5	
rs1185630		rs1882200		rs1152888			rs1185630		rs1882200		rs1152888	
<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<u>First Degree Relative With IBD</u>												
Yes	15/138	11%	37/146	25%	11/206	5%	8/62	13%	18/60	30%	3/64	5%
No	26/224	12%	82/262	31%	20/260	8%	36/224	16%	72/239	31%	14/242	6%
<u>Presence of Extra-intestinal Manifestations</u>												
Yes	19	12%	49	28%	14	8%	15/104	14%	30/110	27%	6/112	5%
No	22/202	11%	70/234	30%	17/228	7%	29/182	16%	60/186	32%	11/194	6%
<u>Age of Disease Onset ^a</u>												
Early	7/94	7%	27/100	27%	3/100	3%	19/138	14%	47/150	31%	11/150	7%
Late	24/204	12%	65/232	28%	19/232	8%	4/60	7%	20/66	30%	2/66	3%

Table 18. *MyD88* mutant allele frequencies for CD and UC populations at the sub-phenotype level. Comparisons of mutant allele frequencies were performed between different subgroups of the same phenotypic variable (i.e. CD patients with a family history of IBD vs. CD patients with no family history of IBD) for the same disease group using Fishers Exact test. Allele frequency is given as: “mutant allele frequency / total number of alleles”. Statistically significant differences in mutant allele frequencies are highlighted in red.

^a Early age of onset in this instance was defined as ≤ 20 years for CD and ≤ 40 for UC. The frequency of UC with onset ≤ 20 years was low (13% of total UC population) and comparisons would have been meaningless. Comparisons for CD with onset ≤ 40 years are described below (see Vienna Classification).

Crohn's Disease								Ulcerative Colitis							
<i>MyD88</i>		<i>MyD88</i>		<i>MyD88</i>				<i>MyD88</i>		<i>MyD88</i>		<i>MyD88</i>			
Intron 3		3'-UTR		3'-UTR				Intron 3		3'-UTR		3'-UTR			
rs4988457		rs7744		rs6853				rs4988457		rs7744		rs6853			
<i>N</i>	%	<i>N</i>	%	<i>N</i>	%			<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
<u>First Degree Relative With IBD</u>															
Yes	11/144	8%	31/144	22%	22/144	15%		1/60	2%	10/62	16%	6/60	10%		
No	13/250	5%	44/252	17%	18/246	7%		16/218	7%	38/240	16%	33/238	14%		
<u>Presence of Extra-intestinal Manifestations</u>															
Yes	10/170	6%	31/168	18%	20/170	12%		9/110	8%	22/110	20%	16/106	15%		
No	14/224	6%	40/228	18%	20/220	9%		8/168	5%	26/192	14%	23/188	12%		
<u>Age of Disease Onset ^a</u>															
Early	9/100	9%	14/96	15%	15/100	15%		8/142	6%	27/148	18%	16/144	11%		
Late	11/228	5%	44/224	20%	19/216	9%		6/62	10%	11/64	17%	10/64	16%		

CD 21/70 [30%] vs. control frequency 24/180 [13%]; $P=0.003$). In contrast, none of the *IRAK-M* mutant alleles appeared to be associated with any of the phenotypic variables outlined in Table 14.

Estimation of Haplotype Frequencies: Haplotype frequencies were estimated using the expectation maximization method as described previously, and were considered for all the phenotypic variables listed in Tables 13 and 14. Estimated haplotype frequencies were compared in a similar manner as in the population-based tests described above. None of the haplotype frequency comparisons were significant for *IRAK-M*. However, one *MyD88* haplotype appeared to have a significantly different distribution between CD patients subdivided on the basis of age of onset (Vienna Classification groups A1 and A2). Estimated haplotype frequencies for the early onset (group A1) and late onset CD (group A2) are shown in Table 19. The global test of the null hypothesis for the comparison between the two age of CD onset groups was rejected ($P=0.049$), suggesting that one or more of the haplotypes have different frequencies in the two groups. Four individual tests were performed to identify haplotypes with different frequencies between the two groups (see Table 19). These showed that the haplotype consisting of the mutant rs7744 (3'-UTR) allele and wildtype rs4988457 (intron 3) and rs6853 (3'-UTR) alleles was significantly more common in the late onset (A2) group than in the early onset (A1) CD group (frequency 0.28 vs. 0.14; $P=0.025$). However, following application of the Bonferroni correction for 4 individual tests (i.e. adjusted significance level $P=0.0125$), this result was not significant. Type 1 error therefore cannot be excluded in this instance.

<i>MyD88</i> Haplotype (rs4988457-rs7744-rs6853)	Early Onset CD n=138 (Group A1)		Late Onset CD n=32 (Group A2)		χ^2	P
	Freq.	s.d.	Freq.	s.d.		
-/-/-	0.73	0.03	0.64	0.06	1.87	0.171
-/+/-	0.14	0.02	0.28	0.06	5.01	0.025
-/-/+	0.06	0.02	0.05	0.03	0.25	0.617
+/-/-	0.01	0.01	0.02	0.02	0.00	0.965

Table 19. Estimated haplotype frequencies for *MyD88* SNPs in early and late onset CD populations (Vienna Classification groups A1 and A2). Haplotypes were estimated from a total population of 170 CD patients. Estimated *MyD88* haplotype frequencies and standard deviations are given as a proportion of the group total. Haplotypes with a significantly different distribution between groups are highlighted in red.

+ Mutant allele, – Wildtype allele.

Family-Based Tests of Association: Family-based tests were performed in the same phenotypic groups. Two PDT analyses showed evidence for *MyD88* mutant allele association in the CD population (Table 20). Firstly, there was over-representation of the *MyD88* intron 3 rs4988457 mutant allele in probands with fistulizing CD when compared to their unaffected siblings in 17 families ($P=0.045$). Secondly, there was over-transmission of the 3'-UTR rs7744 mutant allele from parents to CD-affected individuals in 11 families where the disease process was not complicated by extra-intestinal manifestations ($P=0.039$). However, neither of these analyses were significant following application of the Bonferroni method, using a corrected significance level of $P \leq 0.0167$. The opposite situation was evident for analysis of *MyD88* mutant allele frequencies in UC families with a disease process uncomplicated by the presence of extra-intestinal manifestations. PDT analysis of 11 families showed under-representation of the 3'-UTR rs7744 mutant allele in UC patients with no extra-intestinal manifestations ($P=0.033$). This was not significant following correction for multiple analyses. In a similar manner, the global test of differentiation for *MyD88* SNP haplotypes using the likelihood ratio test was significant for in a cohort of 26 families where the proband had UC uncomplicated by extra-intestinal manifestations ($P=0.044$) (Table 21). Analysis of transmission of individual haplotypes showed under-transmission of the haplotype consisting of the rs7744 mutant allele, and the rs4988457 and rs6853 wildtype alleles in this UC subgroup ($P=0.018$). As with all other family-based analyses this was not significant at the corrected significance level, and type 1 error can therefore not be excluded.

Table 20. Significant Pedigree Disequilibrium Test analyses for *MyD88* sub-phenotypic groups. These analyses appear to show over-transmission of the two *MyD88* SNPs in both CD with a fistulizing behavior pattern and disease uncomplicated by extra-intestinal disease manifestations. The opposite pattern is seen in UC with the absence of extra-intestinal disease, where mutant alleles are significantly under-transmitted.

Disease	Group	SNP	Number of Families	Mutant Allele Frequency				χ^2	P
				Parental Contribution		Discordant Sibs			
				Transmitted	Not Transmitted	Affected	Unaffected		
CD	Fistulizing (B3)	rs4988457 (Intron 3)	17	0	0	5	2	4.00	0.045
	Absence of Extra-intestinal Manifestations	rs7744 (3'-UTR)	11	4	1	5	1	4.26	0.039
UC	Absence of Extra-intestinal Manifestations	rs7744 (3'-UTR)	11	0	2	2	8	4.57	0.033

<i>MyD88</i> Haplotype (rs4988457-rs7744- rs6853)*	Observed Frequency	Expected Frequency	Variance	χ^2	P
-/-/-	44	40.92	2.30	4.13	0.042
-/+/-	6	9.70	2.44	5.61	0.018
-/-/+	1	0.69	0.10	0.93	0.336
+/-/+	1	0.69	0.11	0.87	0.350

Table 21. *MyD88* haplotype frequencies for families where the proband has UC without the presence of extra-intestinal manifestations. Analyses were performed using 25 families using a likelihood ratio test as implemented in the computer program TRANSMIT. Haplotypes with a significantly different distribution between groups are highlighted in red. *Only haplotypes with $\geq 5\%$ frequency are shown.

+ Mutant allele; – Wildtype allele.

Discussion

Three SNPs in both *IRAK-M* and *MyD88* have been characterized in order to assess whether either gene has a role in IBD susceptibility in the study population. Following SNP-specific PCR analysis, candidate gene allele data was analyzed using population- and family-based statistical methodology. Initial results from case-control analyses, where candidate gene allele frequencies were compared between disease and control populations, proved disappointing. No allelic associations were demonstrated for the six characterized candidate gene SNPs in CD, UC or IC disease populations as a whole. This absence of association in these major IBD categories was confirmed by comparison of estimated haplotype frequencies for candidate gene SNPs, and by performing family-based studies. It must therefore be concluded that neither *IRAK-M* nor *MyD88* are involved in IBD susceptibility when considering CD, UC and IC as a whole.

Analysis of candidate genes at the sub-phenotypic level did provide some more promising results. Specifically, analysis of *MyD88* 3'-UTR SNP mutant allele frequencies suggested that this gene might play a role in susceptibility to late onset CD in patients with a family history of IBD. Population based methods showed over-representation of 3'-UTR rs6853 alleles in CD patients with a family history of IBD ($P=0.015$), and over-representation of 3'-UTR rs7744 alleles in CD patients with an age of disease onset over forty years of age ($P=0.005$). Importantly, both these results appeared to remain valid following correction for compounding of type 1 statistical error, making them particularly exciting findings.

Correction for multiple testing has proven an extremely troublesome problem throughout this study, as well as other parts of this dissertation. A prime example of this

problem is given when considering results from estimation of *MyD88* SNP haplotypes. These showed that the haplotype consisting of the mutant rs7744 (3'-UTR) allele and wildtype rs4988457 (intron 3) and rs6853 (3'-UTR) alleles was significantly more common in the late onset (A2) group than in the early onset (A1) CD group ($P=0.025$). On first examination, this result appears to corroborate the findings of population-based association tests i.e. that the 3'-UTR rs7744 mutant allele is associated with late onset CD. However, following application of the Bonferroni correction for 4 individual tests, this result is not significant and type 1 error therefore cannot be excluded.

The problem of multiple testing again arises when one considers results from family-based association tests. These suggested that in addition to late onset CD in patients with a family history of IBD, *MyD88* might be involved in susceptibility to both fistulizing CD and CD with the absence of extra-intestinal manifestations. Specifically, there appeared to be over-transmission of intron 3 rs4988457 and 3'-UTR rs7744 mutant alleles in each of these CD sub-phenotypes. The possible involvement of *MyD88* mutant SNPs in development of extra-intestinal manifestations in IBD was complicated still further when one considers the findings of likelihood ratio analysis in UC families. In this instance there was significant under-transmission of a haplotype containing the mutant rs7744 allele in families where the proband had UC without extra-intestinal disease. However, as was the case with comparisons of the estimated haplotypes described above, correction for multiple testing rendered these apparent associations non-significant, again implying that they probably have occurred as a result of type 1 error. All these should therefore only be viewed as trends and by no means conclusive evidence of the involvement of *MyD88* in susceptibility to these sub-phenotypes. However, before

dismissing these results as type 1 error it is worthwhile considering the size of the study cohort. Analysis at the sub-phenotype level in this study cohort involves consideration of very low number of families. It is therefore somewhat remarkable that any associations were evident on PDT and TRANSMIT analysis given the stringency and robustness of these tests. It will therefore be very interesting to repeat these analyses when more patients have been accrued.

Given all the issues relating to compounding of type 1 error and correction for this serious problem, it appears that the associations of *MyD88* 3'-UTR rs6853 and rs7744 SNPs with late onset CD in patients with a family history of IBD are of particular interest. Despite the apparent association of the two 3'-UTR SNPs following population-based testing and correction for multiple analyses, a number of prominent reservations do exist in proclaiming *MyD88* as the “next CD susceptibility gene”. Firstly, since the sub-phenotype populations contain relatively few subjects, any testing performed in these groups will lack statistical power. Low statistical power interferes with the accuracy of statistical comparisons, and it will therefore be of great interest to see if the associations evident for the two 3'-UTR SNPs remain apparent in a larger cohort. Secondly, one would expect linkage disequilibrium to exist between the two 3'-UTR SNPs since they are in extremely close generic proximity (rs7744 and rs6853 are separated by 348bp). Subsequently, one would expect the same associations to be found for both SNPs. This was not the case, with the finding of association of rs6853 alleles for CD patients with a family history of IBD not being reciprocated with rs7744. The same can be said with the association of rs7744 alleles in CD patients with an age of disease onset over forty years

of age not being reciprocated with rs6853 alleles. These results must therefore be viewed cautiously.

An obvious criticism of this work is the absence of association with mutant 3'-UTR SNP alleles in the unstratified CD population. Indeed, case-control studies were far from significant for both 3'-UTR SNPs for CD versus control comparisons of mutant allele frequencies (rs7744 $P=0.225$, rs6853 $P=0.651$). Studies with the CD susceptibility gene *NOD2* in other populations have shown a relatively strong association with unstratified CD populations and extremely high levels of association at the sub-phenotype level (i.e. with fibrostenosing ileocolic CD) (31;35;36). This criticism is addressed when previous experiences with *NOD2* in this laboratory are considered. Associations were only found for one of the three *NOD2* disease-causing mutations in the unstratified CD cohort in this study population (unpublished data). However, all three of the disease-causing SNPs were strongly associated with fibrostenosing ileocolic CD. This is really a reflection on the size of this cohort, and the power of statistical analyses in such a population. The ability to detect association in a small cohort at the unstratified level is somewhat limited given the likely small effect of any susceptibility gene (including *NOD2*) combined with relatively low subject numbers. However, a particular strength of this study is the depth and quality of phenotype data available for all subjects. All patients are derived from a single surgeons practice and have clinical follow-up of significant duration, thus allowing for consistency of phenotyping and accurate information regarding changes in the course of disease. Furthermore, a single GI pathologist with a particular interest in IBD has reviewed all pathology, again increasing the accuracy and consistency of phenotyping. Therefore, any analysis of genetic data at

the sub-phenotype level will likely be highly accurate. This, taken together with the previous study of *NOD2* in this population, gives strength to the argument that *MyD88* may be involved in susceptibility to late onset CD in patients with a family history of IBD.

Another point of concern is the use of the notoriously conservative Bonferroni method to correct for compounding of type 1 error in this study. Indeed, some have started to question the rationale for using such stringent correction methods in genetic association studies, and have suggested that guarding against a single false positive occurring in genetic association studies is often going to be much too strict and will lead to many missed findings (131). Furthermore, some have suggested that assessing the likelihood of chance association should be considered in the light of the biological plausibility of any observed associations (95;132). A number of possible SNP associations for the regulatory 3'-UTR of *MyD88* have been rejected using the Bonferroni correction method. It may therefore prove worthwhile to re-analyze the current data set by applying a less stringent means of correcting for multiple analyses.

The case of *IRAK-M* is much more clear-cut. No associations were evident in either the unstratified or stratified disease populations, and it is therefore probably true that this gene plays no role in IBD susceptibility in the study population. It could be argued that low sample size could have led to type 2 statistical errors in the instance of *IRAK-M*. This is, however, unlikely to be the case here since there was not even evidence of trends towards association for any *IRAK-M* SNP in any disease group. It may therefore prove prudent to exclude *IRAK-M* from further study as an IBD susceptibility gene. Negative studies such as this are an unfortunate consequence of the candidate gene

approach. When the biological nature of a disease is uncertain, as is the case with IBD, researchers are required to make “educated guesses” as to mechanisms underlying pathogenesis of the disorder. Although *IRAK-M* appeared to be a biologically plausible candidate that is located in an IBD-associated susceptibility locus, it did not turn out to be involved in disease susceptibility. The cause of IBD2 disease association in this population therefore remains unclear.

In summary, preliminary evidence is provided to suggest that the adaptor protein *MyD88* may be involved in susceptibility to late onset CD in patients with a family history of IBD. The next step in this study will be to increase the size of the cohort and thus increase *MyD88* genotype data. If the associations shown here hold true in a larger population it is probable that *MyD88* is indeed an IBD susceptibility gene.

CHAPTER IX

CONCLUDING REMARKS

Summary of Findings

This dissertation presents new, albeit somewhat preliminary, evidence to suggest that MyD88 is involved in susceptibility to late onset Crohn's disease in patients with a family history of IBD. Furthermore, IRAK-M has been conclusively excluded from IBD susceptibility in this population.

The aim of the first series of experiments was to characterize a number of IBD susceptibility loci in the study population. The overall aim here was to improve study population disease characterization by defining peak LOD score STRs from previously described susceptibility loci. A somewhat different approach was taken than has commonly been employed in previous studies. Firstly, a case-control approach was used, in contrast to the affected sib-pair linkage analyses of earlier work. Secondly, distinct IBD phenotypic groups were examined, rather than the clinically heterogeneous 'CD' and 'UC' classification system conventionally employed. Disease associations were observed for the IBD1, IBD2 and IBD5 susceptibility loci in a variety of IBD phenotypes. In addition, the other studied loci from IBD susceptibility in this population.

The results from this study, as well as a previous study (94), formed the basis of the second set of experiments, where a more detailed characterization of the IBD2

susceptibility locus was performed. This locus was chosen ahead of IBD1 since the susceptibility gene within this locus, *NOD2*, had already been defined (26;27). Similarly, IBD5 was not characterized any further since the region of susceptibility has already been narrowed down to a 250kb region of chromosome 5 (118). In contrast, the IBD2 locus remained less well characterized, and by employing the same methods as were used in the previous study, IBD2 was shown to be involved in susceptibility to indeterminate colitis, colonic Crohn's disease, and to some extent ulcerative colitis in the study population.

This detailed characterization of IBD2 formed the background for the next study, where a candidate gene approach was employed with the aim being to define new IBD susceptibility genes. Two genes were chosen for this purpose: *IRAK-M*, an inhibitor of the proinflammatory Toll-like receptor signaling pathway, was chosen on the basis of its genomic location (within IBD2, close to the peak LOD score STR) and biological plausibility as an IBD candidate gene; *MyD88*, an adaptor protein involved in TLR-mediated apoptotic signaling, was chosen on the basis of its biological plausibility and its genomic location within the IBD9 susceptibility locus. It should be noted, however, that the IBD9 locus was not characterized in this population. The first study essentially formed the groundwork for assessing these two candidates as IBD susceptibility genes. First, colonic mucosal expression levels in IBD-affected tissue were quantified for both candidate genes. This was performed to determine whether candidate genes exhibited dysregulation in IBD-affected colonic mucosa when compared to normal colonic mucosa. Second, candidate gene coding and regulatory regions were screened for polymorphisms. Mutation of these regions could have profound effects on candidate gene functionality,

which in turn could lead to the development of IBD. Both candidates were found to be significantly over-expressed in both UC- and CD-affected colonic mucosa. Additionally, seven commonly occurring SNPs were defined within coding or potentially regulatory regions of the two candidate genes. Four of these SNPs were within *IRAK-M* and three were within *MyD88*.

The series of experiments following this screening process involved characterization of six of the seven candidate gene SNPs in the entire study cohort. The seventh SNP, a T→A polymorphism within exon 12 of *IRAK-M*, was not studied any further since it was a synonymous change and did not alter the amino acid sequence of the IRAK-M protein. Population- and family-based statistical methodologies were employed to examine the role of identified candidate gene SNPs in IBD susceptibility. *MyD88* provided by far the more interesting results. Apparent associations were observed with two SNPs within the 3'-UTR of *MyD88* in late onset CD in patients with a family history of IBD. These associations are particularly interesting since 3'-UTRs have been shown to be extremely important in regulating gene expression, and mutations within the 3'-UTRs of several genes have been implicated in the pathogenesis of a number of diseases (128). These associations are however only view these as preliminary evidence of involvement of *MyD88* in IBD susceptibility for a number of reasons. This is primarily because associations were observed in a relatively small sample size, meaning that subsequent analyses lacked statistical power. *IRAK-M*, on the other hand, did not display any evidence of disease association in this population, and this gene is unlikely to be involved in IBD susceptibility in this population.

Directions of Future Research

The most important question raised by this dissertation is whether *MyD88* is really involved in susceptibility to late onset CD in patients with a family history of IBD. As mentioned previously, the statistical analyses used to derive this conclusion lacked statistical power. For example, the total number of patients in the comparative groups which led to the conclusion that the *MyD88* 3'-UTR rs7744 SNP was associated with late onset CD was 125 in the A1 Vienna Classification subgroup, and 35 in the A2 Vienna Classification subgroup. It is for this reason that the primary focus of this study was to define allelic associations, which enabled the use of increased numbers in analyses, as well as decreasing the total number of comparisons. These numbers are small, and despite the apparent level of association being relatively strong in this instance ($P=0.005$), it would be wise to conclude that the overall sample size is too small to be able to confidently state that *MyD88* is an IBD susceptibility gene. In earlier studies involving the IBD susceptibility gene *NOD2* that have employed this type of sub-phenotype analysis, and the smallest size of any subgroup included has been approximately double the size of the A2 Vienna Classification subgroup used here (31;35;36). For the apparent *MyD88* associations presented here to be accepted by a wider scientific community, it is imperative that newly accrued patients are genotyped for the two *MyD88* 3'-UTR SNPs in the future. Indeed the problem of low sample size was apparent in family based studies, which showed association of 3'-UTR mutant SNP alleles with a number of other CD sub-phenotypes. These were however, not significant after correction for multiple analyses. If these associations hold true in a larger population, it will indeed be an

important finding that will have great implications on our understanding of IBD pathogenesis.

The lack of association of *IRAK-M* SNPs with IBD was disappointing, although not completely unexpected. Negative outcomes are inevitable in the majority of candidate gene studies given our lack of understanding of the pathophysiologic mechanisms underlying the development of most complex disorders. The cause of IBD2 association in this population, and consequently the identity of the IBD2 susceptibility gene remain unknown. There are essentially two approaches that can be taken to attempt to identify the IBD2 susceptibility gene(s): 1) to perform further candidate gene studies with the knowledge that most will have negative outcomes; 2) haplotype block mapping of the susceptibility locus. I believe for a number of reasons that the second option is by far the most appropriate approach to take in future studies. Firstly, high throughput genotyping techniques are now widely available and characterization of polymorphisms in an entire population can be achieved relatively efficiently. Secondly, the analysis software available for haplotype block analysis is both powerful and widely available. Mapping of IBD2 could be performed at a relatively low resolution initially, for example by screening ~1kb amplicons for polymorphisms at 100kb intervals within IBD2. In this manner, the region of association within IBD2 could be narrowed down to a much more manageable size to identify potential candidate genes than the 3.05Mbp region used in the current study. This approach does have a number of serious drawbacks however, the most notable of which is that the process is laborious, costly and by no means guarantees quick results. For example, the fine mapping study that defined the IBD1 gene *NOD2* took approximately six years to complete from start to finish (26).

The identification of IBD susceptibility genes is of crucial importance. Definition of such genes will give researchers fresh insights into the pathological basis of IBD and may aid the development of novel therapies to alleviate the often severe symptoms that accompany this disease. The aim of this study has been to define IBD susceptibility genes, and hopefully it may serve as a platform for future efforts to achieve this goal.

REFERENCES

1. Thomson G, Esposito MS. The genetics of complex diseases. *Trends Cell Biol* 1999;9:M17-M20.
2. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-83.
3. Thorsby E. Invited anniversary review: HLA associated diseases. *Hum Immunol* 1997;53:1-11.
4. Thomson G. HLA disease associations: models for the study of complex human genetic disorders. *Crit Rev Clin Lab Sci* 1995;32:183-219.
5. Satsangi J, Welsh KI, Bunce M et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet* 1996;347:1212-17.
6. Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med* 1991;325:928-37.
7. Rhodes JM, Campbell BJ. Inflammation and colorectal cancer: IBD-associated and sporadic cancer compared. *Trends Mol Med* 2002;8:10-16.
8. Price AB. Overlap in the spectrum of non-specific inflammatory bowel disease--'colitis indeterminate'. *J Clin Pathol* 1978;31:567-77.
9. Loftus EV, Jr., Sandborn WJ. Epidemiology of inflammatory bowel disease. *Gastroenterol Clin North Am* 2002;31:1-20.
10. Watts DA, Satsangi J. The genetic jigsaw of inflammatory bowel disease. *Gut* 2002;50 Suppl 3:III31-III36.
11. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ* 1996;312:95-96.
12. Shanahan F. Inflammatory bowel disease: immunodiagnostics, immunotherapeutics, and ecotherapeutics. *Gastroenterology* 2001;120:622-35.
13. Hugot JP, Zouali H, Lesage S, Thomas G. Etiology of the inflammatory bowel diseases. *Int J Colorectal Dis* 1999;14:2-9.

14. Tobin MV, Logan RF, Langman MJ, McConnell RB, Gilmore IT. Cigarette smoking and inflammatory bowel disease. *Gastroenterology* 1987;93:316-21.
15. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;29:990-996.
16. Roth MP, Petersen GM, McElree C, Feldman E, Rotter JI. Geographic origins of Jewish patients with inflammatory bowel disease. *Gastroenterology* 1989;97:900-904.
17. Ahmad T, Satsangi J, McGovern D, Bunce M, Jewell DP. Review article: the genetics of inflammatory bowel disease. *Aliment Pharmacol Ther* 2001;15:731-48.
18. Hugot JP, Laurent-Puig P, Gower-Rousseau C et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;379:821-23.
19. Satsangi J, Parkes M, Louis E et al. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;14:199-202.
20. Hampe J, Schreiber S, Shaw SH et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999;64:808-16.
21. Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* 2000;66:1857-62.
22. Rioux JD, Silverberg MS, Daly MJ et al. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000;66:1863-70.
23. Cho JH, Nicolae DL, Gold LH et al. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci U S A* 1998;95:7502-7.
24. Hampe J, Frenzel H, Mirza MM et al. Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc Natl Acad Sci U S A* 2002;99:321-26.
25. Duerr RH, Barmada MM, Zhang L et al. Evidence for an inflammatory bowel disease locus on chromosome 3p26: linkage, transmission/disequilibrium and partitioning of linkage. *Hum Mol Genet* 2002;11:2599-606.
26. Hugot JP, Chamaillard M, Zouali H et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599-603.

27. Ogura Y, Bonen DK, Inohara N et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;411:603-6.
28. Inohara N, Nunez G. NODs: intracellular proteins involved in inflammation and apoptosis. *Nat Rev Immunol* 2003;3:371-82.
29. Hugot JP, Zouali H, Lesage S. Lessons to be learned from the NOD2 gene in Crohn's disease. *Eur J Gastroenterol Hepatol* 2003;15:593-97.
30. Hampe J, Cuthbert A, Croucher PJ et al. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;357:1925-28.
31. Cuthbert AP, Fisher SA, Mirza MM et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 2002;122:867-74.
32. Gasche C, Alizadeh BZ, Pena AS. Genotype-phenotype correlations: how many disorders constitute inflammatory bowel disease? *Eur J Gastroenterol Hepatol* 2003;15:599-606.
33. Ahmad T, Armuzzi A, Bunce M et al. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;122:854-66.
34. Louis E, Michel V, Hugot JP et al. Early development of stricturing or penetrating pattern in Crohn's disease is influenced by disease location, number of flares, and smoking but not by NOD2/CARD15 genotype. *Gut* 2003;52:552-57.
35. Hampe J, Grebe J, Nikolaus S et al. Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* 2002;359:1661-65.
36. Lesage S, Zouali H, Cezard JP et al. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;70:845-57.
37. Leong RW, Armuzzi A, Ahmad T et al. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Aliment Pharmacol Ther* 2003;17:1465-70.
38. Yamazaki K, Takazoe M, Tanaka T, Kazumori T, Nakamura Y. Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. *J Hum Genet* 2002;47:469-72.
39. Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002;347:417-29.
40. Rath HC, Schultz M, Freitag R et al. Different subsets of enteric bacteria induce and perpetuate experimental colitis in rats and mice. *Infect Immun* 2001;69:2277-85.

41. Zareie M, Singh PK, Irvine EJ, Sherman PM, McKay DM, Perdue MH. Monocyte/macrophage activation by normal bacteria and bacterial products: implications for altered epithelial function in Crohn's disease. *Am J Pathol* 2001;158:1101-9.
42. Monteleone I, Vavassori P, Biancone L, Monteleone G, Pallone F. Immunoregulation in the gut: success and failures in human disease. *Gut* 2002;50 Suppl 3:III60-III64.
43. Fiocchi C. Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* 1998;115:182-205.
44. Pallone F, Fais S, Squarcia O, Biancone L, Pozzilli P, Boirivant M. Activation of peripheral blood and intestinal lamina propria lymphocytes in Crohn's disease. In vivo state of activation and in vitro response to stimulation as defined by the expression of early activation antigens. *Gut* 1987;28:745-53.
45. Swidsinski A, Ladhoff A, Pernthaler A et al. Mucosal flora in inflammatory bowel disease. *Gastroenterology* 2002;122:44-54.
46. Barnes PJ, Karin M. Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. *N Engl J Med* 1997;336:1066-71.
47. Tak PP, Firestein GS. NF-kappaB: a key role in inflammatory diseases. *J Clin Invest* 2001;107:7-11.
48. Neurath MF, Fuss I, Schurmann G et al. Cytokine gene transcription by NF-kappa B family members in patients with inflammatory bowel disease. *Ann N Y Acad Sci* 1998;859:149-59.
49. Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol* 2001;1:135-45.
50. Medzhitov R, Janeway CA, Jr. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 1997;91:295-98.
51. Vasselon T, Detmers PA. Toll receptors: a central element in innate immune responses. *Infect Immun* 2002;70:1033-41.
52. Janeway CA, Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol* 1989;54 Pt 1:1-13.
53. Medzhitov R, Preston-Hurlburt P, Janeway CA, Jr. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* 1997;388:394-97.
54. Kobe B, Deisenhofer J. Proteins with leucine-rich repeats. *Curr Opin Struct Biol* 1995;5:409-16.

55. Akira S, Sato S. Toll-like receptors and their signaling mechanisms. *Scand J Infect Dis* 2003;35:555-62.
56. Akira S. Toll-like receptor signaling. *J Biol Chem* 2003;278:38105-8.
57. Aravind L, Dixit VM, Koonin EV. Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science* 2001;291:1279-84.
58. Kobayashi K, Hernandez LD, Galan JE, Janeway CA, Jr., Medzhitov R, Flavell RA. IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell* 2002;110:191-202.
59. Takeuchi O, Hoshino K, Kawai T et al. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* 1999;11:443-51.
60. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature* 2001;413:732-38.
61. Poltorak A, He X, Smirnova I et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 1998;282:2085-88.
62. Hayashi F, Smith KD, Ozinsky A et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* 2001;410:1099-103.
63. Hemmi H, Takeuchi O, Kawai T et al. A Toll-like receptor recognizes bacterial DNA. *Nature* 2000;408:740-745.
64. Zarembek KA, Godowski PJ. Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol* 2002;168:554-61.
65. Wesche H, Henzel WJ, Shillinglaw W, Li S, Cao Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* 1997;7:837-47.
66. Huang J, Gao X, Li S, Cao Z. Recruitment of IRAK to the interleukin 1 receptor complex requires interleukin 1 receptor accessory protein. *Proc Natl Acad Sci U S A* 1997;94:12829-32.
67. Burns K, Martinon F, Esslinger C et al. MyD88, an adapter protein involved in interleukin-1 signaling. *J Biol Chem* 1998;273:12203-9.
68. Medzhitov R, Preston-Hurlburt P, Kopp E et al. MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol Cell* 1998;2:253-58.

69. Muzio M, Ni J, Feng P, Dixit VM. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* 1997;278:1612-15.
70. Kawai T, Adachi O, Ogawa T, Takeda K, Akira S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* 1999;11:115-22.
71. Aliprantis AO, Yang RB, Weiss DS, Godowski P, Zychlinsky A. The apoptotic signaling pathway activated by Toll-like receptor-2. *EMBO J* 2000;19:3325-36.
72. Kabra NH, Kang C, Hsing LC, Zhang J, Winoto A. T cell-specific FADD-deficient mice: FADD is required for early T cell development. *Proc Natl Acad Sci U S A* 2001;98:6307-12.
73. Savill J. Apoptosis in resolution of inflammation. *J Leukoc Biol* 1997;61:375-80.
74. Cao Z, Henzel WJ, Gao X. IRAK: a kinase associated with the interleukin-1 receptor. *Science* 1996;271:1128-31.
75. Li S, Strelow A, Fontana EJ, Wesche H. IRAK-4: a novel member of the IRAK family with the properties of an IRAK-kinase. *Proc Natl Acad Sci U S A* 2002;99:5567-72.
76. Wesche H, Gao X, Li X, Kirschning CJ, Stark GR, Cao Z. IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. *J Biol Chem* 1999;274:19403-10.
77. Suzuki N, Suzuki S, Duncan GS et al. Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4.
78. Suzuki N, Suzuki S, Yeh W. IRAK-4 as the central TIR signaling mediator in innate immunity. *Trends Immunol* 2002;23:503.
79. Li L, Cousart S, Hu J, McCall CE. Characterization of interleukin-1 receptor-associated kinase in normal and endotoxin-tolerant cells. *J Biol Chem* 2000;275:23340-23345.
80. Yamin TT, Miller DK. The interleukin-1 receptor-associated kinase is degraded by proteasomes following its phosphorylation. *J Biol Chem* 1997;272:21540-21547.
81. Hu J, Jacinto R, McCall C, Li L. Regulation of IL-1 receptor-associated kinases by lipopolysaccharide. *J Immunol* 2002;168:3910-3914.
82. Ziegler-Heitbrock HW. Molecular mechanism in tolerance to lipopolysaccharide. *J Inflamm* 1995;45:13-26.

83. Salkowski CA, Detore G, Franks A, Falk MC, Vogel SN. Pulmonary and hepatic gene expression following cecal ligation and puncture: monophosphoryl lipid A prophylaxis attenuates sepsis-induced cytokine and chemokine expression and neutrophil infiltration. *Infect Immun* 1998;66:3569-78.
84. Nomura F, Akashi S, Sakao Y et al. Cutting edge: endotoxin tolerance in mouse peritoneal macrophages correlates with down-regulation of surface toll-like receptor 4 expression. *J Immunol* 2000;164:3476-79.
85. McCall CE, Grosso-Wilmoth LM, LaRue K, Guzman RN, Cousart SL. Tolerance to endotoxin-induced expression of the interleukin-1 beta gene in blood neutrophils of humans with the sepsis syndrome. *J Clin Invest* 1993;91:853-61.
86. Mak TW, Yeh WC. Immunology: a block at the toll gate. *Nature* 2002;418:835-36.
87. Bonnert TP, Garka KE, Parnet P, Sonoda G, Testa JR, Sims JE. The cloning and characterization of human MyD88: a member of an IL-1 receptor related family. *FEBS Lett* 1997;402:81-84.
88. Hampe J, Lynch NJ, Daniels S et al. Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. *Gut* 2001;48:191-97.
89. Ruckdeschel K, Mannel O, Schrottner P. Divergence of apoptosis-inducing and preventing signals in bacteria-faced macrophages through myeloid differentiation factor 88 and IL-1 receptor-associated kinase members. *J Immunol* 2002;168:4601-11.
90. Jaunin F, Burns K, Tschopp J, Martin TE, Fakan S. Ultrastructural distribution of the death-domain-containing MyD88 protein in HeLa cells. *Exp Cell Res* 1998;243:67-75.
91. Liebermann DA, Hoffman B. Myeloid differentiation (MyD) primary response genes in hematopoiesis. *Blood Cells Mol Dis* 2003;31:213-28.
92. Wang J, Kobayashi M, Han M et al. MyD88 is involved in the signalling pathway for Taxol-induced apoptosis and TNF-alpha expression in human myelomonocytic cells. *Br J Haematol* 2002;118:638-45.
93. Cho JH, Nicolae DL, Ramos R et al. Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Hum Mol Genet* 2000;9:1425-32.
94. Uthoff SM, Crawford NP, Eichenberger MR et al. Association of ulcerative colitis with the inflammatory bowel disease susceptibility locus IBD2 in non-Jewish Caucasians and evidence of genetic heterogeneity among racial and ethnic populations with Crohn disease. *Am J Med Genet* 2002;113:242-49.

95. Daly AK, Day CP. Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol* 2001;52:489-99.
96. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000;405:847-56.
97. Farmer M, Petras RE, Hunt LE, Janosky JE, Galandiuk S. The importance of diagnostic accuracy in colonic inflammatory bowel disease. *Am J Gastroenterol* 2000;95:3184-88.
98. Gasche C, Scholmerich J, Brynskov J et al. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;6:8-15.
99. Rudolph WG, Uthoff SM, McAuliffe TL, Goode ET, Petras RE, Galandiuk S. Indeterminate colitis: the real story. *Dis Colon Rectum* 2002;45:1528-34.
100. Mehta CR, Patel NR. A Network Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables. *Journal of the American Statistical Association* 1983;78:427-34.
101. Fisher RA. Combining independent tests of significance. *American Statistician* 1948;2:30.
102. Schneider, S. Arlequin: A software for population genetics data analysis. Roessli, D. and Excoffier, L. [2.000]. 2003. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
103. Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000;67:146-54.
104. Martin ER, Bass MP, Kaplan NL. Correcting for a potential bias in the pedigree disequilibrium test. *Am J Hum Genet* 2001;68:1065-67.
105. Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999;65:1170-1177.
106. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996;59:983-89.
107. Vector NTI. 2003. Bethesda, MD, Informax.
108. Xiao W, Oefner PJ. Denaturing high-performance liquid chromatography: A review. *Hum Mutat* 2001;17:439-74.
109. Wavemaker. [4.1]. 2002. Omaha, NE, Transgenomic.

110. Lai E. Application of SNP technologies in medicine: lessons learned and future challenges. *Genome Res* 2001;11:927-29.
111. Lee KM, Park SK, Kim SU et al. N-acetyltransferase (NAT1, NAT2) and glutathione S-transferase (GSTM1, GSTT1) polymorphisms in breast cancer. *Cancer Lett* 2003;196:179-86.
112. Inoue N, Tamura K, Kinouchi Y et al. Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 2002;123:86-91.
113. Curran ME, Lau KF, Hampe J et al. Genetic analysis of inflammatory bowel disease in a large European cohort supports linkage to chromosomes 12 and 16. *Gastroenterology* 1998;115:1066-71.
114. Duerr RH, Barmada MM, Zhang L et al. Linkage and association between inflammatory bowel disease and a locus on chromosome 12. *Am J Hum Genet* 1998;63:95-100.
115. Yang H, Plevy SE, Taylor K et al. Linkage of Crohn's disease to the major histocompatibility complex region is detected by multiple non-parametric analyses. *Gut* 1999;44:519-26.
116. Ma Y, Ohmen JD, Li Z et al. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999;5:271-78.
117. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 2001;60:227-37.
118. Rioux JD, Daly MJ, Silverberg MS et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;29:223-28.
119. Armuzzi A, Ahmad T, Ling KL et al. Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut* 2003;52:1133-39.
120. Orchard TR, Satsangi J, Van Heel D, Jewell DP. Genetics of inflammatory bowel disease: a reappraisal. *Scand J Immunol* 2000;51:10-17.
121. Silverberg MS, Daly MJ, Moskovitz DN et al. Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies. *Gut* 2001;49:773-76.
122. Kaplan N, Morris R. Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. *Genet Epidemiol* 2001;20:432-57.
123. Parkes M, Barmada MM, Satsangi J, Weeks DE, Jewell DP, Duerr RH. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* 2000;67:1605-10.

124. Hamlin PJ, Komolmit P, Bransfield K et al. Identification of multiple candidate genes for IBD susceptibility using high-density transcript mapping in the IBD2 locus on chromosome 12q. *Gastroenterology* 1999;117:1029-31.
125. Moffatt MF, Traherne JA, Abecasis GR, Cookson WO. Single nucleotide polymorphism and linkage disequilibrium within the TCR alpha/delta locus. *Hum Mol Genet* 2000;9:1011-19.
126. Tumer Z, Croucher PJ, Jensen LR et al. Genomic structure, chromosome mapping and expression analysis of the human AVIL gene, and its exclusion as a candidate for locus for inflammatory bowel disease at 12q13-14 (IBD2). *Gene* 2002;288:179-85.
127. van Heel DA, Carey AH, Jewell DP. Identification of novel polymorphisms in the beta7 integrin gene: family-based association studies in inflammatory bowel disease. *Genes Immun* 2001;2:455-60.
128. Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med* 2000;6:637-41.
129. Fedorova L, Fedorov A. Introns in gene evolution. *Genetica* 2003;118:123-31.
130. Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 2003;28:215-20.
131. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440-9445.
132. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ* 1998;316:1236-38.

CURRICULUM VITAE

NIGEL CRAWFORD B.Sc. (Hons), M.B. Ch.B.

DATE OF BIRTH: 5-29-1974

PLACE OF BIRTH: Newton Abbot, United Kingdom.

HOME ADDRESS: 2730 Brownsboro Road #149
Louisville, KY 40206.

OFFICE ADDRESS: Price Institute of Surgical Research,
MDR Bldg Rm 314,
University of Louisville, Louisville, KY 40202.

TELEPHONE: Office: (502) 852-0659
Fax: (502) 852-1256
Home: (502) 893-9973

E-MAIL ADDRESS: nigelcrawford@hotmail.com

EDUCATION:

B.Sc. (Hons) Molecular Biology 2:1 University of Liverpool, Liverpool, United Kingdom.	1995- 1996
M.B. Ch.B. with Commendation, University of Liverpool, Liverpool, United Kingdom.	1992- 1998
Ph.D. Candidate. Department of Physiology and Biophysics. University of Louisville, Louisville, KY. Expected degree conferral date: May 2004.	2001- 2004

B. Sc. (HONS.) THESIS:	The Molecular Biology of Resistance to Alkylating Agents in <i>Streptomyces Coelicolor</i> A3(2).	1995-1996
MEDICAL/SURGICAL INTERNSHIP:	University Hospital Aintree, Liverpool, United Kingdom.	1998-1999
SURGICAL RESIDENCY POSTS: (6 Month Appointments)	University Hospital Aintree, Liverpool, United Kingdom.	1999-2000
	Princess Margaret Hospital, Swindon, United Kingdom.	2000
ACADEMIC APPOINTMENTS:	Anatomy Prosector, University of Oxford, Oxford, United Kingdom.	2000
	Price Fellow in Surgical Research University of Louisville Louisville, Kentucky.	2001 -Present
JOURNAL REVIEWER	<i>American Journal of Surgery</i> Ad hoc reviewer.	2002
AWARDS:	Steele - Pearce Award Highest chemistry A-level result in region (West Devon).	1992

PEER-REVIEWED PUBLICATIONS

Crawford NPS, Colliver DW, Galandiuk S. Tumor markers: utility in management. J Surg Oncol (2003) 84:239-248.

Uthoff SMS, Crawford NPS, Eichenberger MR, Hamilton CJ, Petras RE, Martin ER, Galandiuk S Association of ulcerative colitis with the inflammatory bowel disease susceptibility locus IBD2 in non-Jewish Caucasians and evidence of genetic heterogeneity among racial and ethnic populations with Crohn's disease. Am J Med Genetics (2002) 113:242-249.

Crawford NPS, Colliver DW, Funke AA, Kelley S, Cobbs GA, Petras RE, Galandiuk S. Characterization of Genotype-Phenotype Relationships and Stratification by *NOD2* Variant Genotype for Inflammatory Bowel Disease Susceptibility Loci Using Multiple Short Tandem Repeat Genetic Markers. Hum. Mut. (Submitted).

Crawford NPS, Eichenberger MR, Colliver DW, Lewis RK, Cobbs GA, Petras RE, Galandiuk S. Evaluation of *Slc11a1* (formerly *NRAMP1*) as an Inflammatory Bowel Disease Candidate Gene. Inflammatory Bowel Disease Journal (Submitted).

Crawford NPS, Uthoff SMS, Colliver DW, Funke AA, Petras RE, Cobbs GA, Galandiuk S. Characterization of Genotype-Phenotype Relationships and Stratification by *NOD2* Variant Genotype within the Inflammatory Bowel Disease Susceptibility Locus IBD2. (In preparation).

Crawford NPS, Eichenberger MR, Colliver DW, Kolodko V, Galandiuk S. Characterization of *Nod2* disease-causing mutations in a small, predominantly rural population confirms Crohn's disease genotype-phenotype correlations. (In preparation).

Crawford NPS, Eichenberger MR, Colliver DW, Kolodko V, Galandiuk S. Is there a Role for Toll-Like Receptor Pathway Components in Susceptibility to Inflammatory Bowel Disease? (In preparation)

PUBLISHED ABSTRACTS

Crawford N, Eichenberger MR, McAuliffe TM, Galandiuk S. Regulatory genes of NF- κ B and their effect on IBD. Am J Hum Genet 73(5 suppl) 526.

Crawford N, Uthoff S, Eichenberger MR, Cobbs GA, Petras RE, Martin ER, Galandiuk S (2003). Characterization of genotype-phenotype correlations show that the IBD2 susceptibility locus is associated with colonic Crohn's disease and ulcerative colitis. Gastroenterol 124(No 4, Suppl 1): A48

Crawford N, Eichenberger M, Funke A, Kelley S, Petras R, Galandiuk S (2002). Microsatellite markers help direct surgery for colonic IBD. Gastroenterol 122:A217.

RESEARCH GRANTS

S. Galandiuk, P.I.; N. Crawford, Investigator (2001-2002)
Jewish Hospital Foundation, Rudd Surgical Teaching Endowment
"Genetic analysis of inflammatory bowel disease susceptibility loci within the Kentuckiana population"

S. Galandiuk, P.I.; N. Crawford, Investigator (2001-2002)
James Graham Brown Cancer Center Microarray Pilot Grant Program
"The Use of Microarray Technology as a Diagnostic Tool in Colonic Inflammatory Bowel Disease"

S. Galandiuk, P.I. – Pending
N. Crawford. Research fellow (20% effort)
Philip Morris External Research Program
"Cigarette smoke exposure and IBD"

S. Galandiuk, P.I. – Pending
N. Crawford. Research fellow (20% effort)
NIH/PAR-01-056
"Cigarette smoke exposure and IBD"

S. Galandiuk, P.I. – Resubmission: March 2004
N. Crawford. Research fellow (30% effort)
N.I.H./PA-01-129
"Regulatory genes of NF κ B and their effect on IBD"

S. Galandiuk, P.I. – Resubmission: March 2004
N. Crawford. Research fellow (30% effort)
N.I.H./PA-03-003
"Gene expression in UC dysplasia/cancer"

PRESENTATIONS AT SCIENTIFIC MEETINGS

Regulatory genes of NF- κ B and their effect on IBD. Poster presentation: American Society of Human Genetics, Los Angeles, CA, November 2003.

Characterization of genotype-phenotype correlations show that the IBD2 susceptibility locus is associated with colonic Crohn's disease and ulcerative colitis. Oral presentation: American Gastroenterological Association, Orlando, FL, May 2003.

The use of microsatellite markers as predictors of carcinoma and dysplasia in ulcerative colitis. Poster presentation: World Congress of Epidemiology, Montreal, ON, August 2002.

Microsatellite markers help decide surgery in colonic IBD. Poster presentation: American Gastroenterological Association, San Francisco, CA, May 2002.

Genetic analysis of inflammatory bowel disease susceptibility loci within the Kentuckian population. Oral presentation: Kentucky Chapter of the American College of Surgeons, Louisville, KY, April 2002.